CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 124, No. 5 November 2008

ACOUSTICAL NEWS-USA 2647

USA Meeting Calendar 2650

ACOUSTICAL NEWS-STANDARDS 2656

Standards Meeting Calendar 2656

ADVANCED-DEGREE DISSERTATION ABSTRACTS 2664

BOOK REVIEWS 2665

REVIEWS OF ACOUSTICAL PATENTS 2667

TUTORIAL REVIEW [10]

Methods for reconstructing acoustic quantities based on acoustic pressure measurements — Sean F. Wu — 2680

LETTERS TO THE EDITOR

Linearly filtered estimation of the time-domain Green's function from measurements of ambient noise (L) — S. A. Albahrani, M. R. Frater, E. H. Huntington — 2699

Comment on "Acoustic chaos in a duct with two separate sound sources" [J. Acoust. Soc. Am. 110, 120–126 (2001)] (L) — A. A. Castrejón-Pita, J. R. Castrejón-Pita, G. Huelsz, A. Sarmiento-Galán — 2702

Comment on "Cepstrum pitch determination" [J. Acoust. Soc. Am. 41, 293–309 (1967)] (L) — Arturo Camacho — 2706

Individualized head-related transfer functions based on population grouping (L) — Song Xu, Zhizhong Li, Gavriel Salvendy — 2708

A re-evaluation of the relation between physiological channel interaction and electrode pitch ranking in cochlear implants (L) — Michelle L. Hughes — 2711

GENERAL LINEAR ACOUSTICS [20]

Acoustical scattering by radially stratified scatterers — Liang-Wu Cai, José Sánchez-Dehesa — 2715

NONLINEAR ACOUSTICS [25]

Acoustic radiation torque on an irregularly shaped scatterer in an arbitrary sound field — Zongwei Fan, Deqing Mei, Keji Yang, Zichen Chen — 2727

A resonance shift prediction based on the Boltzmann–Ehrenfest principle for cylindrical cavities with a rigid sphere — Arturo O. Santillan, Vicente Cutanda-Henríquez — 2733

(Continued)

CONTENTS—*Continued from preceding page*

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

A theoretical treatment of the long range propagation of impulsive signals under strongly ducted nocturnal conditions	Roger Waxler, Kenneth E. Gilbert, Carrick Talmadge	2742
Near- to far-field characteristics of acoustic radiation through plug flow jets	G. Gabard	2755
Transitional wall pressure fluctuations on axisymmetric bodies	Chinsuk Hong, Ku-Kyun Shin, Jae-Jin Jeon, Sang-Youn Kim	2767

UNDERWATER SOUND [30]

A new equation for the accurate calculation of sound speed in all oceans	Claude C. Leroy, Stephen P. Robinson, Mike J. Goldsmith	2774
Observations of clustering inside oceanic bubble clouds and the effect on short-range acoustic propagation	Thomas C. Weber	2783
Volumetric (tomographic) three-dimensional geoacoustic inversion in shallow water	A. Tolstoy	2793
Comparison of measured and modeled temporal coherence of sound near 75 Hz and 3683 km in the Pacific Ocean	John L. Spiesberger	2805
Temporal coherence after multiple forward scattering through random three-dimensional inhomogeneities in an ocean waveguide	Tianrun Chen, Purnima Ratilal, Nicholas C. Makris	2812
Passive acoustic detection of deep-diving beaked whales	Walter M. X. Zimmer, John Harwood, Peter L. Tyack, Mark P. Johnson, Peter T. Madsen	2823
Automatic detection of marine mammals using information entropy	Christine Erbe, Andrew R. King	2833
Waveguide invariant broadband target detection and reverberation estimation	Ryan Goldhahn, Granger Hickman, Jeffrey Krolik	2841
An active acoustic tripwire for simultaneous detection and localization of multiple underwater intruders	Thomas Folegot, Giovanna Martinelli, Piero Guerrini, J. Mark Stevenson	2852

ULTRASONICS, QUANTUM ACOUSTICS, AND PHYSICAL EFFECTS OF SOUND [35]

Analytical time-domain Green's functions for power-law media	James F. Kelly, Robert J. McGough, Mark M. Meerschaert	2861
Scattering of the fundamental shear horizontal mode in a plate when incident at a through crack aligned in the propagation direction of the mode	M. Ratassepp, M. J. S. Lowe, P. Cawley, A. Klauson	2873
Influence of material viscoelasticity on the scattering of guided waves by defects	Mihai V. Predoi, Michel Castaings, Ludovic Moreau	2883
Scattering of the fundamental shear horizontal guided wave by a part-thickness crack in an isotropic plate	P. Rajagopal, M. J. S. Lowe	2895

STRUCTURAL ACOUSTICS AND VIBRATION [40]

Scattering of a Bessel beam by a sphere: II. Helicoidal case and spherical shell example	Philip L. Marston	2905

NOISE: ITS EFFECTS AND CONTROL [50]

Generating nonstationary multisensor signals under a spatial coherence constraint	Emanuël A. P. Habets, Israel Cohen, Sharon Gannot	2911
Membrane covered duct lining for high-frequency noise attenuation: Prediction using a Chebyshev collocation method	Lixi Huang	2918
Annoyance responses to stable and changing aircraft noise exposure	Mark Brink, Katja E. Wirth, Christoph Schierz, Georg Thomann, Georg Bauer	2930

(Continued)

CONTENTS—*Continued from preceding page*

ARCHITECTURAL ACOUSTICS [55]

A transient boundary element method model of Schroeder diffuser scattering using well mouth impedance	Jonathan A. Hargreaves, Trevor J. Cox	2942
Analytical prediction of the breakout noise from a rectangular cavity with one compliant wall	B. Venkatesham, Mayank Tiwari, M. L. Munjal	2952

ACOUSTIC SIGNAL PROCESSING [60]

Signal denoising and ultrasonic flaw detection via overcomplete and sparse representations	Guang-Ming Zhang, David M. Harvey, Derek R. Braden	2963
Joint deconvolution and classification with applications to passive acoustic underwater multipath	Hyrum S. Anderson, Maya R. Gupta	2973

PHYSIOLOGICAL ACOUSTICS [64]

Transient evoked otoacoustic emission latency and estimates of cochlear tuning in preterm neonates	Arturo Moleti, Renata Sisto, Alessia Paglialonga, Federica Sibella, Lucien Anteunis, Marta Parazzini, Gabriella Tognola	2984
Transient evoked otoacoustic emission input/output function and cochlear reflectivity: Experiment and model	Renata Sisto, Arturo Moleti	2995
Effects of low-frequency biasing on spontaneous otoacoustic emissions: Frequency modulation	Lin Bian	3009
Auditory brainstem responses to a chirp stimulus designed from derived-band latencies in normal-hearing subjects	Claus Elberling, Manuel Don	3022
Representation of the spectral dominance region of pitch in the steady-state temporal discharge patterns of cochlear nucleus units	William P. Shofner	3038
Objective threshold estimation and measurement of the residual background noise in auditory evoked potentials of goldfish	Jianqiang Xiao, Christopher B. Braun	3053

PSYCHOLOGICAL ACOUSTICS [66]

The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms	Nicole Marrone, Christine R. Mason, Gerald Kidd, Jr.	3064
Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation	Etienne Gaudrain, Nicolas Grimault, Eric W. Healy, Jean-Christophe Béra	3076
Discrimination of interaural temporal disparities conveyed by high-frequency sinusoidally amplitude-modulated tones and high-frequency transposed tones: Effects of spectrally flanking noises	Leslie R. Bernstein, Constantine Trahiotis	3088
Can the binaural system extract fine-structure interaural time differences from noncorresponding frequency channels?	David A. Magezi, Katrin Krumbholz	3095
Binaural model for artificial spatial sound localization based on interaural time delays and movements of the interaural axis	Laurent Kneip, Claude Baumann	3108
Relative influence of interaural time and intensity differences on lateralization is modulated by attention to one or the other cue	Albert-Georg Lang, Axel Buchner	3120
Discrimination and identification of azimuth using spectral shape	Daniel E. Shub, Suzanne P. Carr, Yunmi Kong, H. Steven Colburn	3132
Current-level discrimination and spectral profile analysis in multi-channel electrical stimulation	Matthew J. Goupell, Bernhard Laback, Piotr Majdak, Wolf-Dieter Baumgartner	3142

(Continued)

CONTENTS—*Continued from preceding page*

SPEECH PRODUCTION [70]

Speech production variability in fricatives of children and adults: Results of functional data analysis — Laura L. Koenig, Jorge C. Lucero, Elizabeth Perlman — 3158

Temporal organization of English clear and conversational speech — Rajka Smiljanić, Ann R. Bradlow — 3171

A linear model of acoustic-to-facial mapping: Model parameters, data set size, and generalization across speakers — Matthew S. Craig, Pascal van Lieshout, Willy Wong — 3183

Perception and production of /r/ allophones improve with hearing from a cochlear implant — Melanie L. Matthies, Frank H. Guenther, Margaret Denny, Joseph S. Perkell, Ellen Burton, Jennell Vick, Harlan Lane, Mark Tiede, Majid Zandipour — 3191

SPEECH PERCEPTION [71]

Identification of frequency-shifted vowels — Peter F. Assmann, Terrance M. Nearey — 3203

A model for multitalker speech perception — Soundararajan Srinivasan, DeLiang Wang — 3213

Study of an automated procedure for a Dutch sentence test for the measurement of the speech reception threshold in noise — Hayo Terband, Rob Drullman — 3225

Identification of Mandarin tones by English-speaking musicians and nonmusicians — Chao-Yang Lee, Tsun-Hui Hung — 3235

The role of temporal cues in word identification by younger and older adults: Effects of sentence context — Sandra Gordon-Salant, Grace Yeni-Komshian, Peter Fitzgibbons — 3249

SPEECH PROCESSING AND COMMUNICATION SYSTEMS [72]

Speech production modifications produced by competing talkers, babble, and stationary noise — Youyi Lu, Martin Cooke — 3261

MUSIC AND MUSICAL INSTRUMENTS [75]

Experimental investigation of an American five-string banjo — Laurie A. Stephey, Thomas R. Moore — 3276

Interaction of reed and acoustic resonator in clarinetlike systems — Fabrice Silva, Jean Kergomard, Christophe Vergez, Joël Gilbert — 3284

Influence of a constriction in the near field of the vocal folds: Physical modeling and experimental validation — Lucie Bailly, Xavier Pelorson, Nathalie Henrich, Nicolas Ruty — 3296

BIOACOUSTICS [80]

Frequency dependence of backscatter from thin, oblique, finite-length cylinders measured with a focused transducer-with applications in cancellous bone — Keith A. Wear, Gerald R. Harris — 3309

Bowhead whale springtime song off West Greenland — Kathleen M. Stafford, Sue E. Moore, Kristin L. Laidre, M. P. Heide-Jørgensen — 3315

Beamwidth control and angular target detection in an echolocating bottlenose dolphin (*Tursiops truncatus*) — Patrick W. Moore, Lois A. Dankiewicz, Dorian S. Houser — 3324

JASA EXPRESS LETTERS

On the exponent in the power law for the attenuation at low frequencies in sandy sediments — William M. Carey, Allan D. Pierce, Richard E. Evans, Jason D. Holmes — EL271

(Continued)

CONTENTS—*Continued from preceding page*

A study of high frequency ultrasound scattering from non-nucleated biological specimens	Omar Falou, Ralph E. Baddour, George Nathanael, Gregory J. Czarnota, J. Carl Kumaradas, Michael C. Kolios	EL278
Spectral attenuation of sound in dilute suspensions with nonlinear particle relaxation	Max Kandula	EL284
Parkinson's disease and the effect of lexical factors on vowel articulation	Peter J. Watson, Benjamin Munson	EL291
High-frequency dispersion from viscous drag at the grain-grain contact in water-saturated sand	Nicholas P. Chotiros, Marcia J. Isakson	EL296
Psychometric functions for hybrid difference discrimination/ increment detection tasks	Daniel Shepherd, Michael J. Hautus, Hamish Hutchinson	EL302
Low-frequency attenuation of acoustic waves in sandy/silty marine sediments	Allan D. Pierce, William M. Carey	EL308
Preliminary mapping of void fractions and sound speeds in gassy marine sediments from subbottom profiles	T. G. Leighton, G. B. N. Robb	EL313
The effect of signal-temporal uncertainty on detection in bursts of noise or a random-frequency complex	Angela Yarnell Bonino, Lori J. Leibold	EL321
Linear frequency modulation reverberation suppression using difference of singular values	Hyung Soo Lee, Kyun Kyung Lee	EL328

CUMULATIVE AUTHOR INDEX 3341

On the exponent in the power law for the attenuation at low frequencies in sandy sediments

William M. Carey, Allan D. Pierce, and Richard E. Evans
Mechanical Engineering Department, College of Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215
wcarey@bu.edu, adp@bu.edu, richard.evans.01@snet.net

Jason D. Holmes
BBN Technologies, 10 Moulton Street, Cambridge, Massachusetts 02318
jholmes@bbn.com

Abstract: Shallow water transmission loss measurements yield intrinsic attenuation estimates for acoustic waves in the underlying sediment, with results that are consistent with attenuation being proportional to frequency raised to a power n, with n between 1.6 and 1.87. Plausible theory suggests that n should be identically 2. The discrepancy can be explained because the inverse analysis inferences were made with the neglect of an additional attenuation mechanism where generated lower velocity shear waves carry energy downwards out of the waveguide. The shear wave effect has a weaker dependence on frequency than the intrinsic attenuation, so the apparent exponent is shifted downward.

© 2008 Acoustical Society of America
PACS numbers: 43.30.Bp, 43.30.Cq, 43.30.Ma [JL]
Date Received: June 14, 2008 **Date Accepted:** July 28, 2008

1. Introduction

Accurate estimates of sound transmission loss and time spread are required in the design and performance prediction of sonar systems from the low (50 Hz) to the mid- (1–5 kHz) frequencies. Calculation of shallow water sound transmission has been shown to be accurate when geoacoustic profiles derived from geophysical measurements with site-specific frequency and depth dependent attenuation characteristics are used. In particular, coastal margins and seas formed by depositional layers of sands and silt are of naval interest. Shallow water areas with water depths between 40 and 200 m and sandy bottoms that have a compressional wave speed greater than that of the water are classified as fast or critical angle bottoms, whereas the depositional fans of major rivers composed of fine silts and mud are slow bottoms. Zhou *et al.*[1] recognized that in these sandy shallow water areas nonlinear frequency dependence of the acoustic attenuation in the upper sedimentary layer is required to explain many aspects of shallow sound propagation. That is, the intrinsic attenuation can be described by a nonlinear power law of the form

$$\alpha(f) = \alpha(f_0)(f/f_0)^n. \tag{1}$$

$\alpha(f_0)$ [NP/m or dB/m] is the intrinsic attenuation and f_0 is a reference frequency.

The above conclusion, based on the experimental results of Zhou *et al.* and the work of Stoll,[2] proved instrumental in shallow water sonar system demonstrations and confirmatory sound transmission investigations.[3–5] In addition, a summary[6] of the results from experiments conducted under known environmental and bottom conditions showed that accurate calculation of narrow- and broadband sound transmission in fast waveguides requires a nonlinear frequency dependence ($n=1.8$) for attenuation in the upper sediment layer for frequencies between 50 Hz and 1 kHz. This summary was consistent with a larger selection of results including nonunique inversions reviewed by Zhou and Zhang[7] to yield $\alpha(1\text{ kHz})=0.34$ and $n\cong 1.84$.

However, a direct consequence of the Biot theory[8] is that sandy sediments should have a quadratic dependence, $n=2$.

2. A simplified version of the Biot Model

Biot's equations derived from the fundamental Lagrange–Euler equations with the forces per unit volume exerted on the fluid by the solid matter and the equal and opposite force exerted by the fluid can be simplified[8] as

$$\rho\partial^2\Phi/\partial t^2 - \lambda\nabla^2\Phi - (D/b\lambda^2)\partial^3\Phi/\partial t^3 = 0. \tag{2}$$

Here ρ is the density, λ is a Lamé constant, b is the equivalent dashpot constant per unit volume, and D is a parameter that characterizes the lack of perfect coupling between the fluid and the solid motions including both inertial and elastic constants. The interesting character of this equation is the third time derivative. The above equation is, in fact, the standard wave equation with this added term. If we consider plane wave propagation

$$\Phi \rightarrow \Phi_0 \exp(-i\omega t + ikx), \tag{3}$$

then upon substitution the following dispersion equation is obtained:

$$-\rho\omega^2 + \lambda k^2 - i\omega^3(D^2/\rho b\lambda^2) = 0. \tag{4}$$

The complex wave number is readily obtained as

$$k = (\omega/(\lambda/\rho)^{1/2})[1 + i\omega(D^2/\rho b\lambda^2)]^{1/2}. \tag{5}$$

When the frequency is low and the imaginary term is much less than one, a binominal expansion yields the complex wave number

$$k = \omega/c + i\alpha \quad \text{with} \quad c = (\lambda/\rho)^{1/2}, \quad \alpha(\omega) = (1/2(\rho\lambda)^{1/2})(D^2/b\lambda^2)\omega^2. \tag{6}$$

Thus in the low-frequency limit the attenuation is proportional to frequency squared, $n=2$. The attenuation is inversely proportional to b and viscosity. The factor $\tau = D^2/\rho\, b\lambda^{\ 2}$ is referred to as a time constant and when substituted in the first equation yields

$$\nabla^2\Phi \ -(1/c^2)\partial^2\Phi/\partial t^2 - = -(\tau/c^2)\partial^3\Phi/\partial t^3. \tag{7}$$

The order of magnitude estimates can be made from experimental results when one considers the attenuation $\alpha(\omega) = \tau\,\omega^2/2c$. Rogers (see Ref. 6: in Ref. 6) yielded $3.4\times10^{-6} < \tau < 4.8\times10^{-6}$ s and Beebe (see Ref. 3 in Ref. 6) yielded $22\times10^{-6} < \tau < 37\times10^{-6}$ s. Beebe's result has a larger reference attenuation constant while Roger's result was based on a constant depth dependent attenuation; these factors may explain the time constant differences. Zhou and Zhang's[7] empirical relationship at frequencies less than several kilohertz is $\alpha = 0.34(f/1\ \text{kHz})^{1.84}$ consistent with Hamilton's (see Ref. 16 in Ref. 6) The corresponding range is $2.7\times10^{-6} < \tau < 3.73\times10^{-6}$ s. Since the time constant is inversely proportional to viscosity and since viscosity divided by density has dimensions of characteristic length squared per unit time, then geometrically similar sediments having the same composition and porosities should yield a time constant that scales inversely with characteristic length squared, and the grain size squared.

3. Propagation effects

The propagation of sound in a depth dependent waveguide with absorption is an improper Sturm–Liouville problem.[8] The depth-separated wave equation is

$$\rho\frac{d}{dz}\left(\frac{1}{\rho}\frac{dZ}{dz}\right) + \left(\frac{\omega^2}{c^2} + i\frac{\tau}{c^2}\omega^3 - k^2\right)Z = 0. \tag{8}$$

The Rayleigh quotient for this problem may be written in terms of the (depth dependent) complex constant and the modal eigenfunction Z_n as

$$k_n^2 = \frac{\int \left[\frac{\omega^2}{c^2} + i\frac{\tau}{c^2}\omega^3 \right] \frac{Z_n^2}{\rho} dz - \int \frac{1}{\rho} \left(\frac{dZ_n}{dz} \right)^2 dz}{\int \frac{Z_n^2}{\rho} dz}. \quad (9)$$

The complex modal wave number with modal attenuation α_n and phase speed $V_{\mathrm{ph},n}$ is

$$k_n^2 = \left(\frac{\omega}{V_{\mathrm{ph},n}} + i\alpha_n \right)^2. \quad (10)$$

A perturbation solution for the modal coefficients that was originally developed by Kornhauser and Raney (1955) (see Ref. 11 in Ref. 9) and Ingenito (1973) (Ref. 12 in Ref. 9) showed that the modal attenuation coefficient is

$$\alpha_n = V_{\mathrm{ph},n} \frac{\int \frac{\alpha(\omega)}{\rho c} Z_n^2 dz}{\int \frac{Z_n^2}{\rho} dz}. \quad (11)$$

The integral is taken over the waveguide and the result is a good approximation if the modal functions determined in the absence of attenuation are used. The modal attenuation coefficient depends on the integral of the product of intrinsic attenuation and the modal eigenfunction squared. The low-order low-frequency eigenfunctions penetrate deeper in the sediment than the higher-order higher-frequency functions; consequently the ratio decreases with increasing frequency. In general for a sandy bottom that has depth dependent sound speed, attenuation, density, and porosity profiles, the measured attenuation of sound in the water portion of the waveguide will have frequency dependence less than quadratic. However, numerical field calculations that use the range-depth dependencies of the bottom should be capable of determining the correct intrinsic frequency dependence when sufficient experimental sound transmission and environmental data are available; the reason being that the depth dependent attenuation effects are incorporated in the attenuation profiles and are reflected in the calculations.

4. The Nantucket sound results

Numerical studies have shown factors such as rough surface scattering and anomalous sound speed perturbations due to internal waves cannot explain the inferred frequency dependence of intrinsic attenuation. To further examine this finding, sound transmission experiments were conducted in the Nantucket Sound[9,10] where surface roughness and internal waves were not factors in the determination of the transmission loss. The transmission loss was measured for frequencies between 220 and 1228 Hz in a nominal water depth of 13–14 m to ranges of 3600 m. The propagation path was over a fairly constant depth sandy-silty bottom protected on the north and west by Tuckernuck Shoal and on the south and east by Nantucket (see Fig. 1 in Ref. 9). The sediment in the upper layers of the bottom was 97% sand, 3% silt, 49% porosity, and density 1.7 g/cm^3. Sound speed, attenuation, and density profiles as a function of depth were determined based on these average bottom properties. The sound speed in the water was measured using sound speed, temperature, and depth probe on the ship as well as one on board the vehicle to be 1467 ± 0.25 m/s and constant over depth. The sound speed profile in the sediment increased from 1550 to 1800 m/s with a gradient consistent with Hamilton's empirical relations. Density was considered to be a constant over depth to the acoustic basement, as the small gradient predicted by theory would not be influential at the frequencies employed in this experiment (see Fig. 2 in Ref. 9; see Ref. 10, pp. 179–186).

Comparison of measured range averaged transmission loss with fast field, normal mode, and parabolic equation calculations using simplified geoacoustic profiles was excellent

(Ref. 10, pp. 187–189) when a nonlinear frequency dependent attenuation was employed. The reference attenuation was determined by the 1228 Hz transmission measurements. An iterative comparison between range dependent parabolic equation calculations yielded a most probable value of the power exponent, $n=1.87$, and the 95% confidence interval $1.74<n<2.04$ with $0.261<\alpha(f_0=1228)<0.273$ in the first 5.5 m of the sediment. While this value is consistent with Biot at the upper end of the confidence interval, the $n=1.87$ value provides the least bias between calculated and measured estimates.

The measurement of the bottom and subbottom with frequencies of 33 and 200 kHz indicated that the roughness, layering, and range dependence were not significant in this experiment. The remaining factor to be considered was shear wave conversion. An analytical treatment of a waveguide with a water layer (ρ,c) over a sediment (ρ_b,c_b) with a slow shear wave speed $(c_s<c<c_b)$ shows that the rate of energy conversion to shear waves could explain the apparent additional attenuation observed, $\alpha_a(f)=\alpha_{\mathrm{swc}}+\alpha_i(f)$. Here α_{swc} describes the removal of energy from the water borne field by shear waves. At higher frequencies $\alpha_a(f)\approx\alpha_i(f)$ while at lower frequencies $\alpha_a(f)\approx\alpha_{\mathrm{swc}}$; thus, the observed effective attenuation is determined by a power law relationship between these asymptotes with a less than quadratic dependence. This effective shear wave attenuation can be shown to be proportional to the cube of the shear wave speed, c_s^3.[11] Hamilton (see Ref. 16 in Ref. 6) and Stoll[2] reported shear wave speeds between 100 and 300 m/s for sandy sediments with porosities of 40%–50%. Hastrup (Ref. 12, pp. 121–127) reported empirical relationships that relative speed ratios $1.0\leqslant c_b/c\leqslant 1.079$ yield shear wave speeds $184\leqslant c_s\leqslant 400$ m/s. While the actual value of the shear wave speed and gradients in the first 5.5 m of sediments are largely unknown, these estimates indicate that the effective attenuation due to shear wave conversion can vary by a factor of 64. The importance of shear wave conversion in bottom reflection has been treated by Hastrup[12] while the depth dependent properties were investigated by Chapman and Godin.[13]

5. Numerical results

In the analysis of the transmission measurements an iterative comparison was made with calculations with a liquid isospeed layer over a fluid half-space with constant properties. The attenuation was allowed to have the power law frequency dependence, Eq. (1), with the resulting inferred value of $n=1.7$. Although this simplest of geoacoustic models provides a good comparison with our measurements and insight with respect to the modal variations it does not capture the requisite geoacoustic characteristics. With this in mind calculations were performed on this problem with a solid bottom and a frequency dependent attenuation.

A normal mode code based on Ivansson and Karasalo's study[14] was used to compute transmission loss for this case at four frequencies: 125, 250, 500, and 1000 Hz. The source and receiver depths are both 6 m. The water layer was 13 m deep with a sound speed of 1467.5 m/s. The bottom half-space has a compressional speed of 1620 m/s, a density of 1.7 gm/cm^3, and compressional wave attenuation $\alpha_i(f_0)(f/f_0)^2$ with $\alpha_i(f_0=1\ \mathrm{kHz})=0.30$ dB/m(0.49 dB/λ). The sediment half-space was treated as a solid, $c_s=300$ m/s, or a fluid, $c_s=0$ m/s.

The transmission loss, with cylindrical spreading removed, was intensity averaged over the range interval 2–10 km using a 0.2 km sliding window. The intensity averaged transmission loss was fitted with a straight line. Table 1 has the intercepts and slopes of the least square fits for the four frequencies and the solid and fluid bottoms.

The range averaged transmission loss, $\langle \mathrm{TL}(f,\ r)\rangle_{0.2\ \mathrm{km}}$, is shown in Fig. 1. These computational results show the importance of shear wave conversion at the lower frequencies and negligible effects at the higher frequencies. That is at a frequency of 1 kHz the solid and fluid bottom (dashed line) cases produce indistinguishable results; while at the frequencies of 125 and 250 Hz a large difference is observed. One must keep in mind that the attenuation due to shear wave conversion, α_{swc}, is proportional to c_s^3 and consequently these results could change dramatically with the value chosen and its variation with depth. These results show that basic inversions that employ compressional effects may produce estimates of $n<2$.

Table 1. Intercepts (dB) and slopes (dB/km) of least squares fit to the intensity averaged transmission loss, with cylindrical spreading removed.

Freq. (Hz)	Intercept (dB) (solid)	Slope (dB/km) (solid)	Intercept (dB) (fluid)	Slope (dB/km) (fluid)
1000	14.7	0.43	14.6	0.39
500	12.9	0.75	12.9	0.50
250	11.2	1.34	11.4	0.66
125	11.7	1.44	11.8	0.67

An example of this effect is shown in Fig. 2 where a comparison of the $\langle \mathrm{TL}(f,r)\rangle_{0.2\ \mathrm{km}}$ is shown for two cases. The first case, solid line, is for a solid bottom with $c_s=300$ m/s and a frequency dependence of $n=2$. The second case is for a fluid bottom, dashed line.

The attenuation $\alpha(f_0=1\ \mathrm{kHz})$ (dB/m) was chosen based on the measured results and common to all these calculations. The iterative comparison technique was employed between the solid and the fluid case to determine the best fit, a value of $n=1.6$. The consequence of this analysis is that inversions that employ fluid bottoms will even under the simplest of conditions yield a frequency dependence determined by the relative importance of α_{swc} and $\alpha_{\ i}(f)$. Gradients, realistic layering, and other depth variations further complicate this situation;[13] however, these affects will not affect the qualitative role of shear wave energy conversion on measured attenuation. Nevertheless it is possible to quantify the attenuation, $\alpha_a=\alpha_{\mathrm{wc}}(f,z)+\alpha_i(f,z)$, and to represent the attenuation as $\alpha_a=\alpha_i(f_0)(f/f_0)^m$ where m is a specific power exponent to compensate for shear in fluid bottom calculations.

6. Summary and conclusions

The nonlinear frequency dependence of the attenuation inferred from sound transmission measurements in shallow water waveguides with sedimentary sand bottoms has been shown from 50 Hz to 1 kHz to be of the form $\alpha(f)=\alpha(f_0)(f/f_0)^n$ with $\alpha(f_0=1\ \mathrm{kHz})\approx 0.3-0.4$ dB/m and $1.6\leqslant n\leqslant 1.87$. Simplified Biot theory shows that this dependence should be quadratic with $n=2$. Calculations incorporating scattering, layering, and compressional wave speed gradients were found not to explain the less than quadratic dependence. However, calculations with geoacoustic profiles including shear show that sound transmission attenuates as a result of shear

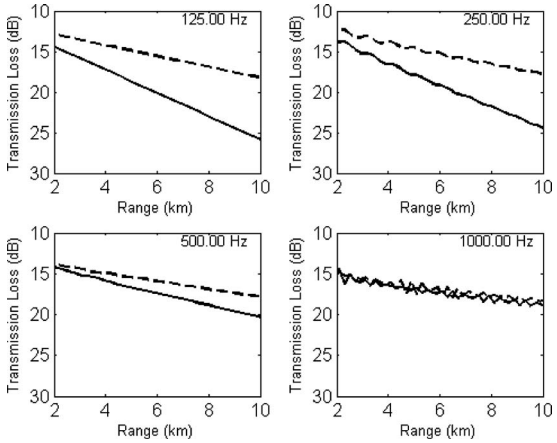


Fig. 1. A comparison of cylindrically corrected range averaged transmission loss, $\langle \mathrm{TL}(f,r)\rangle_{0.2\ \mathrm{km}}$, for frequencies of 125, 250, 500, and 1000 Hz. Calculations are presented for solid bottom, $c_s=300$ m/s (solid line) and fluid bottom, $c_s=0$ m/s (dashed line).

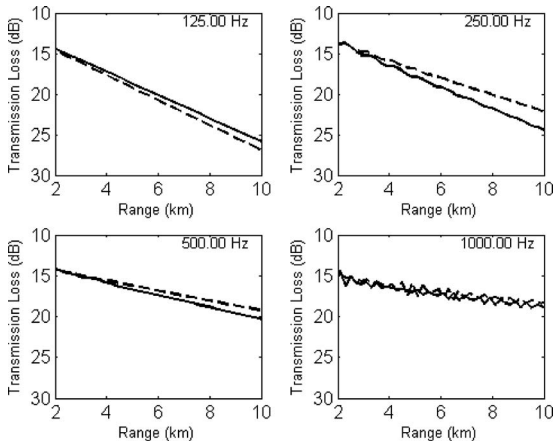


Fig. 2. A comparison of cylindrically corrected range averaged transmission loss, $\langle \mathrm{TL}(f,r)\rangle_{0.2\ \mathrm{km}}$, for frequencies of 125, 250, 500, and 1000 Hz. Calculations are presented for the solid bottom (c_s=300 m/s, n=2) case (solid line) and the fluid bottom (c_s=0 m/s, n=1.6) case (dashed line).

wave conversion of acoustic energy at the water sediment interface and attenuation in the sediment. Since the intrinsic sediment attenuation, $\alpha_i(f)$, and the apparent attenuation due to shear wave conversion, α_{swc}, are small, the effective attenuation can be $\alpha_a(f) \approx \alpha_{\mathrm{swc}} + \alpha_i(f)$. If the removal of energy from the field due to shear wave conversion is comparable to the intrinsic attenuation at the lower frequencies but less than the intrinsic attenuation at the higher frequencies, the inferred frequency dependence can be less than quadratic. Calculations with fluid sediments can account for this effect by the use of a site-specific m factor that accounts for the loss of energy from the compressible field due to shear wave conversion. Further the knowledge of $c_s(f,z)$ is required and may necessitate new or improved methods to experimentally determine this property of sandy sediments.

Acknowledgments

The authors would like to thank the support of the Ocean Acoustics Program of ONR.

References and links

[1]J. X. Zhou, X. Z. Zhang, P. H. Rogers, and J. Jarzynski, "Geoacoustic parameters in a stratified sea bottom from shallow-water acoustic propagation," J. Acoust. Soc. Am. **82**, 2068–2074 (1987).
[2]R. Stoll, *Sediment Acoustics* (Springer-Verlag, New York, 1989).
[3]R. Evans and W. Carey, "Frequency dependence of sediment attenuation in two low-frequency shallow-water acoustic experimental data sets," IEEE J. Ocean. Eng. **23**, 439–447 (1998).
[4]I. Rozenfeld, W. Carey, P. Cable, and W. Siegmann, "Modeling and analysis of sound transmission in the Strait of Korea," IEEE J. Ocean. Eng., **26**(4), 809–820 (2001).
[5]S. M. Dediu, W. M. Carey, and W. L. Siegmann, "Statistical analysis of sound transmission results on the New Jersey continental slope," J. Acoust. Soc. Am. **122**, EL23–EL28 (2007).
[6]J. D. Holmes, W. M. Carey, S. M. Dediu, and W. L. Siegmann, "Nonlinear frequency-dependent attenuation in sandy sediments," J. Acoust. Soc. Am. **121**, EL218–EL222 (2007); see also J. D. Holmes, S. Dediu, and W. M. Carey, "Nonlinear frequency dependent attenuation in sandy sediments," Boston University Technical Report, March 2005.
[7]J. X. Zhou and X. Z. Zhang, "Nonlinear frequency dependence of the effective sea bottom acoustic attenuation from low-frequency field measurements in shallow water," J. Acoust. Soc. Am. **117**, 2494 (2005).
[8]A. D. Pierce, W. M. Carey, and M. Zampoli, "Low-frequency attenuation of sound in marine sediments," in *Proceedings of Oceans 2005*, Brest, France (IEEE, New York, OES, 2005).
[9]J. D. Holmes, W. M. Carey, and J. F. Lynch, "Results from an autonomous underwater vehicle towed hydrophone array experiment in Nantucket Sound," J. Acoust. Soc. Am. **120**(2), EL15–EL21 (2006).
[10]J. D. Holmes, "Investigation of ocean acoustics using autonomous instrumentation to quantify the water-sediment boundary properties," Ph.D. thesis, Boston University, Boston, MA, 2007.
[11]A. O. Williams and R. K. Eby, "Acoustic attenuation in a liquid layer over a "slow" viscoelastic solid," J. Acoust. Soc. Am. **34**, 836–843 (1962).

[12]O. F. Hastrup, "Acoustic bottom reflectivity," Technical Report No. SR-115, SACLANT, Undersea Research Centre, La Spezia, Italy, 1986.
[13]D. M. Chapman and O. A. Godin, "Dispersion of interface waves in sediments with power-law shear speed profiles. II. Experimental observations and seismo-acoustic inversions," J. Acoust. Soc. Am. **110**, 1908–1916 (2001).
[14]S. Ivansson and I. Karasalo, "Computation of modal wavenumbers using an adaptive winding-number integral method with error control," J. Sound Vib. **161**, 173–180 (1993).

A study of high frequency ultrasound scattering from non-nucleated biological specimens

Omar Falou
Department of Electrical and Computer Engineering, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada
ofalou@ryerson.ca

Ralph E. Baddour and George Nathanael
Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada
rbaddour@uhnres.utoronto.ca,
gnathana@uwo.ca

Gregory J. Czarnota
Department of Radiation Oncology, Toronto Sunnybrook Regional Cancer Centre, Toronto, Ontario M4N 3M5, Canada
gregory.czarnota@sunnybrook.ca

J. Carl Kumaradas and Michael C. Kolios
Department of Physics, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada
ckumarad@ryerson.ca,
mkolios@ryerson.ca

Abstract: The high frequency backscatter from cells with a nucleus to cell volume ratio of 0.50 cannot be adequately modeled as a homogeneous sphere. It was hypothesized that the cytoplasm of such cells is of fluid nature. This work attempts to model the ultrasound backscatter (10–62 MHz) from some non-nucleated biological specimens. This was done by measuring the backscatter response from individual sea urchin oocytes and comparing it to theoretical predictions in both the time and frequency domains. A good agreement was found between the experimental and theoretical results suggesting that the non-nucleated oocytes are of fluid nature.

© 2008 Acoustical Society of America
PACS numbers: 43.80.Cs, 43.80.Gx [Charles Church]
Date Received: June 17, 2008 **Date Accepted:** July 30, 2008

1. Introduction

When compared to clinical ultrasound imaging (1–10 MHz), high frequency ultrasound imaging (20–60 MHz) is more sensitive to cell structure and cell spatial distribution changes.[1,2] Recent publications have demonstrated that high frequency ultrasound has the potential of being used for tumor classification.[3] Others have shown that high frequency ultrasound can be used to detect structural and physical changes in cell ensembles during apoptosis.[4] Ultrasonic backscatter from cell ensembles treated with the chemotherapeutic drug cisplatin, which induces apoptosis (apoptosis is described by Kerr *et al.*[5]), increased the ultrasound backscatter signal amplitude by 9–13 dB and induced changes in the frequency dependence of backscatter.[4,6] Theoretical models of ultrasound scattering at the cellular level are needed in order to develop methods for using ultrasound backscatter measurements to classify tumors or determine their response to treatment. The development of these models requires an understanding of the mechanical properties of components of a cell such as the nucleus and the cytoplasm.

Baddour *et al.*[7] performed successful measurements of high frequency (10–65 MHz) backscatter from single cells. A recent study by the same group showed that for prostate carci-

 © 2008 Acoustical Society of America

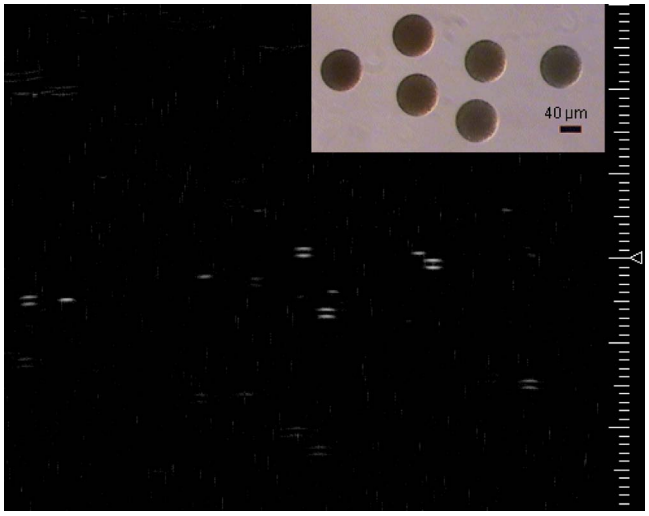


Fig. 1. (Color online) B scan of a sparse suspension of sea urchin oocytes in seawater at 40 MHz (smallest scale =100 μm). The triangle on the right hand side of the image indicates the location of the transducer focus. Inset is an optical microscopy of sea urchin oocytes.

noma (PC-3) cells whose nucleus to cell volume ratio is 0.33, the backscatter response could be modeled as a fluid sphere.[8] However, the same study found that for human acute myeloid leukemia (OCI-AML-5) cells whose nucleus to cell volume ratio equals to 0.50, the backscatter response was not modeled as well by a fluid sphere. Baddour *et al.* hypothesized that the cytoplasm is of fluid nature whereas the nucleus possesses elastic properties, giving rise to this discrepancy.[8] This work attempts to model the ultrasound backscatter (10–62 MHz) from some non-nucleated biological specimens through the measurement and comparison of the backscatter response from sea urchin oocytes to theoretical predictions from a fluid sphere.

2. Methods

2.1 Calculation of backscatter transfer function

A sparse suspension of oocytes from *Strongylocentrotus purpuratus* (purple sea urchin) were prepared in artificial seawater (ρ=1025 kg/m^3, c=1527 m/s).[9] The oocytes were obtained by shedding live female urchins into artificial seawater with 0.5*M* KCL. They are mainly composed of chromatin and yolk proteins. These oocytes were selected because their spherical shape and narrow size distribution, as shown in the inset of Fig. 1 (mean oocyte diameter of approximately 75 μm, standard deviation=2.2 μm).

Data acquisition was performed using a VisualSonics VS40B ultrasound imaging device (VisualSonics Inc., Canada). Three transducers, with different resonant frequencies, f-numbers, and focal lengths, were employed. Table 1 summarizes the properties of the three transducers. Data from the −6 dB bandwidth of each spectrum of the incident pulse (as measured by the reflection off a quartz plate) were used in the analysis, giving an overall analysis range of 10–62 MHz. Ten independent acquisitions of 75, 100, and 125 linearly separated (150 μm spacing) raw rf echo signals were performed using the 20, 40, and 80 MHz transducers, respectively. Since oocytes centered in the focal region of a given transducer lead to a higher amplitude echo opposed to oocytes that are located off the transducer axis, any rf scan lines whose maximum was less than 90% of the overall maximum value from all rf lines were discarded, assuming they did not contain an oocyte completely in the focal region of that scan.

Table 1. Properties of the transducers used in the experiments

Transducer	f-number	Focal length (mm)	Excitation frequency (MHz)	−6 dB bandwidth (MHz)
20 MHz polyvinylidene fluoride	2.35	20	19	10–28
40 MHz polyvinylidene fluoride	3.00	9	40	21–51
80 MHz lithium niobate	3.00	6	55[a]	40–62

[a]The 80 MHz transducer was pulsed at 55 MHz in order to expand the frequency band for this study to 62 MHz. The performance of the transducer was satisfactory.

A Hamming window of 2 μs width, centered at the maximum value of each rf line, was applied to all remaining lines in order to partially remove abnormal scattering patterns in some lines due to the presence of more than one oocyte in the focal region (resolution volume) of the scan. Furthermore, visual inspection was performed to eliminate any lines which exhibited these patterns and were not removed by the previous step. For each transducer, the remaining rf lines (of which there were between 5 and 23) were translated to align the midpoint of the two largest positive peaks and then averaged to obtain a single rf line corresponding to the backscatter from an individual "average" oocyte. From this, a backscatter transfer function $\mathrm{BSTF}_{\mathrm{expr}}(\omega)$ was calculated:

$$\mathrm{BSTF}_{\mathrm{expr}}(\omega) = \frac{R_{\mathrm{expr}}(\omega)}{R_{\mathrm{ref}}(\omega)}, \tag{1}$$

where $R_{\mathrm{expr}}(\omega)$ is the Fourier transform of the average backscatter signal (within the transducer's depth of field). $R_{\mathrm{ref}}(\omega)$ is the Fourier transform of the average reference signal. Reference signals were obtained using the specular reflection from a flat polished SiO_2 crystal (Edmund Industrial Optics Inc., part 43424; $\rho = 2200$ kg/m^3, $c = 5720$ m/s) placed at the transducer focus in seawater. The average reference signal was obtained by superimposing and averaging 32 independent acquisitions of rf echo signals for each transducer. The values of the $|\mathrm{BSTF}|^2$ are presented in the form of spectral plots expressed in decibels relative to the backscatter intensity from the reference (dB_r). These were compared with the theoretical backscatter frequency responses calculated for a fluid sphere using the Anderson model,[10] with a range of densities and sound speeds for the scattering sphere. A least squares analysis was used to determine the density and speed of sound for the fluid sphere that best agreed with the experimental response in the spectral domain. This was performed by minimizing the sum of the squares of the difference between theoretical and experimental values. Densities from 1140 to 1260 kg/m^3 and speeds of sound from 1540 to 1600 m/s were tried.

2.2 Time domain signal reconstruction

Previous studies of scattering from individual cells[7,8] focused on the analysis of experimental results in the frequency domain through the BSTF. The analysis of data in time domain provides additional insight into scattering behavior of objects, such as ringing patterns and ring-down time. In this work, experimental and theoretical backscatter signals from sea urchin oocytes are presented and compared in both time and frequency domains. Given the experimental transmitted pulse and the theoretical backscatter frequency responses, the constructed time domain echo is given by

$$r_{\mathrm{theor}}(t) = \mathcal{F}^{-1}\{R_{\mathrm{ref}}(\omega)\mathrm{BSTF}_{\mathrm{theor}}(\omega)\}, \tag{2}$$

where, $\mathcal{F}^{-1}\{\,\}$ is the inverse Fourier transform and $\mathrm{BSTF}_{\mathrm{theor}}(\omega)$ is the scattered echo of a homogeneous sphere insonified by a plane wave of unit amplitude.[10,7]

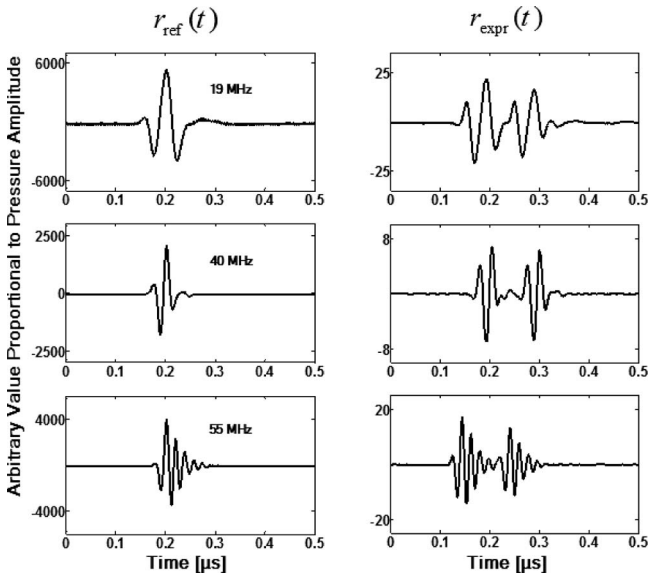


Fig. 2. Average backscatter echoes from individual sea urchin oocytes and their corresponding incident pulses from three transducers: 20, 40, and 80 MHz excited at 19, 40, and 55 MHz, respectively.

3. Results

Figure 1 shows the B scan of sparse suspension of sea urchin oocytes in seawater at 40 MHz. Each pair of bright bands corresponds to the backscatter from a single oocyte. For each transducer, the average backscatter echoes from individual sea urchin oocytes (using the method described previously) and their corresponding average incident pulses are presented in Fig. 2. The measured backscatter frequency responses within a range that corresponds to the −6 dB bandwidths of each transducer are plotted in Fig. 3. A theoretical density of 1198 kg/m^3 and a speed of sound of 1573 m/s for the oocyte were found to have the best fit to experimental data. The theoretical frequency backscatter response of a fluid sphere (d=75 μm, c=1573 m/s, ρ =1198 kg/m^3) is also shown in Fig. 3. The average experimental and reconstructed signals from individual sea urchins oocytes from the 40 MHz transducer (which produced the most distinct double bright bands per oocyte) are presented in Fig. 4.

4. Discussion and Conclusions

There is a very good agreement in the locations of maxima and minima, as shown in Fig. 3. The difference has been measured to be less than 1% "on average" between the theoretical fre-

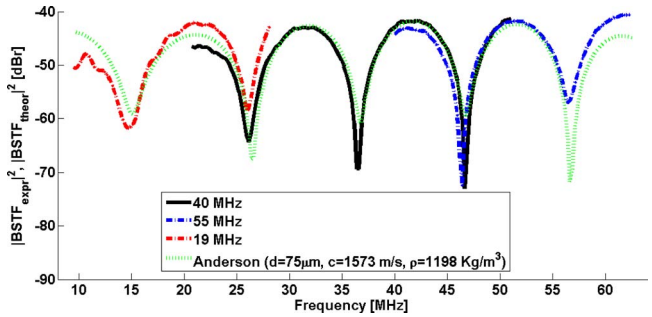


Fig. 3. (Color online) Theoretical (fluid sphere model parameters: d=75 μm, c=1573 m/s, ρ=1198 kg/m^3) and measured backscatter frequency responses of single sea urchin oocytes in seawater subject to three incident pulses from three transducers: 20, 40, and 80 MHz excited at 19, 40, and 55 MHz, respectively.

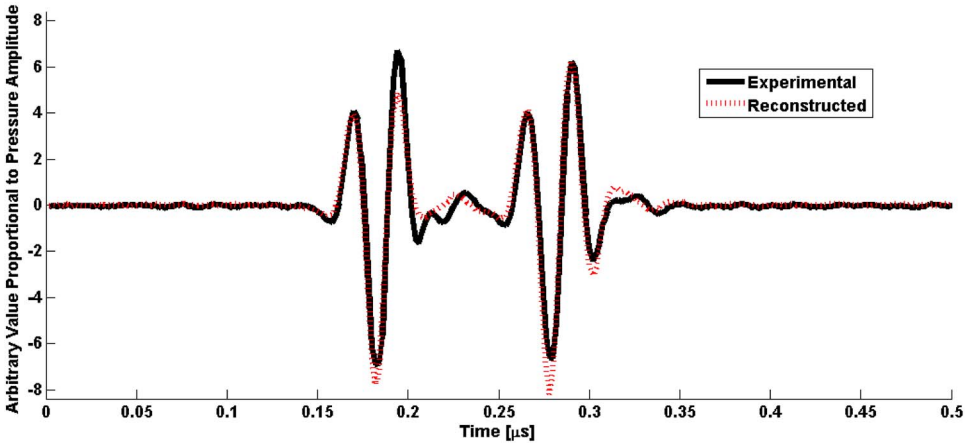


Fig. 4. (Color online) Average experimental and reconstructed (given the experimental transmitted pulse and the theoretical backscatter frequency responses) echoes from individual sea urchin oocytes at 40 MHz. Note that the reflection off the front and the back surface is similar to the incident pulse and there is no phase reversal at the back surface.

quency response of a fluid sphere and the measured backscatter frequency response of individual sea urchin oocytes. The presence of many ripples in the experimental data collected using the 20 MHz transducer may be due to relatively low number of rf lines left (five of them) after discarding inappropriate scans (as described in Sec. 2). This suggests that most of the oocytes in the solution were not located within the −6 dB focal volume of the transducer in the particular experiment. The overall agreement shown in Fig. 3 suggests that sea urchin oocytes can be best modeled as fluid sphere. These findings are not surprising since sea urchin oocytes have no membrane delimited nucleus and have a spherical shape (inset of Fig. 1).

The incident pulses and backscatter echoes for each transducer (Fig. 2) reveal that sea urchins are very weak scatterers [peak of echoes from quartz plate were ~300 times (−49.5 dB) larger than that of the backscatter echoes]. No ringing pattern was observed which indicates the absence of detectable shear waves inside the oocytes, a further indication of their fluidlike nature. Scattering seems to occur off the front and back surfaces of the sea urchin oocytes, as shown in the B-scan image (Fig. 1). The backscatter echo is composed of two pulses which can be easily distinguished for the 40 MHz transducer. The peak-to-peak separation of the latter is 0.096 μs which corresponds to a distance of 151 μm (the distance was found by multiplying the peak-to-peak separation by the speed of sound in the oocyte) or approximately twice the mean diameter of the oocytes. These findings further strengthen the argument that the oocytes can be modeled as fluid spheres.

A good agreement was found in the time domain between the experimental and reconstructed echoes from individual oocytes, as shown in Fig. 4. The maxima/minima that are larger than ±1 a.u. align temporally with no measureable error and their values agree to an average error of 24%. In macroscopic models of wave propagation, a phase inversion occurs in the reflection at the boundary between two media when a wave travels from a region with higher acoustic impedance to lower impedance. The experimental rf data (Fig. 4) do not exhibit this phase reversal at the back surface of the oocyte (where the wave travels from higher to lower impedance). These results suggest that, although scattering mainly occurs at the surrounding medium-oocyte interfaces, a complete scattering model, such as that of Anderson, is needed to predict the scattered field from spherical objects.

In conclusion, the backscatter response from sea urchin oocytes were measured and compared to theoretical predictions for a fluid sphere (Anderson model). Very good agreement was found between the experimental and theoretical results in the frequency domain (difference less than 1% on average) suggesting that the backscatter response from a sea urchin oocyte is

best modeled as a fluid sphere. This work confirms the fluid behavior of the non-nucleated biological specimens used in this study at high ultrasound frequencies. It also establishes an objective methodology to determine the appropriate theoretical model of the properties of individual cells as ultrasound scatterers. Future work will include the use of this methodology to test a hypothesis that acute myeloid leukemia cells, whose nucleus to cell volume ratio equals to 0.5, are best modeled as an elastic nucleus surrounded by a fluid cytoplasm. Computational methods (such as the finite-element analysis) will be used to solve the theoretical model.

Acknowledgments

The authors would like to thank Mr. Zenon Harley and Dr. Homayoun Vaziri for their expertise and guidance and providing the sea urchin oocytes used in this study. This project was funded by the Canadian Institutes of Health Research (Grant No. 79447). This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program awarded to M.C.K. O.F. is the recipient of an Ontario Graduate Scholarship. Funding to purchase the equipment was provided by the Canada Foundation for Innovation, the Ontario Ministry of Research and Innovation and Ryerson University.

References and Links

[1]J. W. Hunt, A. E. Worthington, A. Xuan, M. C. Kolios, G. J. Czarnota, and M. D. Sherar, "A model based upon pseudo regular spacing of cells combined with the randomisation of the nuclei can explain the significant changes in high-frequency ultrasound signals during apoptosis," Ultrasound Med. Biol. **28**, 217–226 (2002).

[2]A. S. Tunis, G. J. Czarnota, A. Giles, M. D. Sherar, J. W. Hunt, and M. C. Kolios, "Monitoring structural changes in cells with high-frequency ultrasound signal statistics," Ultrasound Med. Biol. **31**, 1041–1049 (2005).

[3]M. L. Oelze, W. D. O'Brien, Jr., J. P. Blue, and J. F. Zachary, "Differentiation and characterization of rat mammary fibroadenomas and 4T1 mouse carcinomas using quantitative ultrasound imaging," IEEE Trans. Med. Imaging **23**, 764–771 (2004).

[4]G. J. Czarnota, M. C. Kolios, H. Vaziri, S. Benchimol, F. P. Ottensmeyer, M. D. Sherar, and J. W. Hunt, "Ultrasonic biomicroscopy of viable, dead and apoptotic cells," Ultrasound Med. Biol. **23**, 961–965 (1997).

[5]J. F. R. Kerr, A. H. Wyllie, and A. R. Currie, "Apoptosis—Basic biological phenomenon with wide-ranging implications in tissue kinetics," Br. J. Cancer **26**, 239–257 (1972).

[6]M. C. Kolios, G. J. Czarnota, M. Lee, J. W. Hunt, and M. D. Sherar, "Ultrasonic spectral parameter characterization of apoptosis," Ultrasound Med. Biol. **28**, 589–597 (2002).

[7]R. E. Baddour, M. D. Sherar, J. W. Hunt, G. J. Czarnota, and M. C. Kolios, "High-frequency ultrasound scattering from microspheres and single cells," J. Acoust. Soc. Am. **117**, 934–943 (2005).

[8]R. E. Baddour and M. C. Kolios, "The fluid and elastic nature of nucleated cells: Implications from the cellular backscatter response," J. Acoust. Soc. Am. **121**, EL16–EL22 (2007).

[9]K. V. MacKenzie, "Nine-term equation for sound speed in the oceans," J. Acoust. Soc. Am. **70**, 807–812 (1981).

[10]V. C. Anderson, "Sound scattering from a fluid sphere," J. Acoust. Soc. Am. **22**, 426–431 (1950).

Spectral attenuation of sound in dilute suspensions with nonlinear particle relaxation

Max Kandula
ASRC Aerospace, NASA Kennedy Space Center, Florida 32899
max.kandula-1@ksc.nasa.gov

Abstract: Previous studies on the sound attenuation in particle-laden flows under Stokesian drag and conduction-controlled heat transfer have been extended to accommodate the nonlinear drag and heat transfer. It has been shown that for large particle-to-fluid density ratio, the particle Reynolds number bears a cubic relationship with $\omega\tau_d$ (where ω is the circular frequency and τ_d is the Stokesian particle relaxation time). This dependence leads to the existence of a peak value in the linear absorption coefficient occurring at a finite value of $\omega\tau_d$. Comparison of the predictions with the test data for the spectral attenuation of sound with water injection in a perfectly expanded supersonic air jet shows a satisfactory trend of the theory accounting for nonlinear particle relaxation processes.

© 2008 Acoustical Society of America
PACS numbers: 43.20.Hq, 43.50.Nm, 43.50.Gf [AN]
Date Received: June 23, 2008 **Date Accepted:** August 7, 2008

1. Introduction

Sound attenuation in fluids, representing the dissipation of acoustic energy from a sound wave, occurs through a number of physical processes involving molecular viscosity, thermal conductivity, and other dissipative or relaxation processes.[1–4] In all these absorption mechanisms, acoustic energy is converted into thermal energy, and other mechanisms deflect or scatter acoustic energy.[1] Stokes[5] developed the first successful theory of sound absorption due to the effect of molecular viscosity of the fluid (internal friction). Helmholtz[6] theoretically investigated the effect of viscosity on the sound attenuation in a circular tube, considering the effect of friction near solid boundaries. Kirchhoff[7] theoretically treated sound attenuation due to the viscosity and heat conduction effects in an unbounded fluid medium and in narrow tubes. The so-called *classical sound absorption* in fluids includes Stokes' viscous contribution and Kirchhoff's thermal conduction contribution to attenuation in unbounded fluids.

When a fluid contains inhomogeneities such as suspended particles (solid particles, drops, and bubbles) additional viscous and heat conduction losses occur in the immediate neighborhood of the suspended particles.[1–3,8] A comprehensive review of the physics and scientific history of acoustic interactions with particulate mixtures is provided by Challis *et al.*[9] Referring to Fig. 1, the acoustic intensity I of a plane wave propagating through an absorbing medium is expressed by

$$I = I_0 e^{-\alpha_i x}, \tag{1}$$

where x is the distance traversed, I_0 the intensity at $x=0$, and α_i the intensity attenuation coefficient for the medium. The quantity α_i depends on viscosity, thermal conductivity, and other factors such as molecular relaxation.

Sound propagation in aerosols and fog has been studied experimentally and theoretically by several investigators. Sewell,[10] in his pioneering work, theoretically considered the attenuation of sound in a viscous medium containing suspended cylindrical or spherical particles (obstacles) with perfectly rough surfaces. With the aid of a scattering formulation (theory), Sewell predicted the attenuation of sound by rigid particles suspended in a gas, assuming that the particles are immovable. Epstein,[11] in a theoretical treatment of the attenuation of sound by spherical particles suspended in liquids or gases, derived and extended Sewell's result

 © 2008 Acoustical Society of America

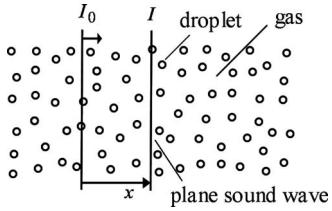


Fig. 1. Attenuation of a plane sound wave in a gas-droplet mixture.

by permitting the particles to move. Epstein and Carhart[12] additionally considered heat conduction effects and found that for fogs the effects of viscosity and heat conduction are both important and approximately additive. Their formulation was extended by Allegra and Hawley[13] to include liquid-liquid as well as liquid-solid systems (with a different model for each case).

The effect of transport processes on the attenuation and dispersion of sound in aerosols has also been reported by Soo,[14] accounting for nonstationary effects including the Basset history (effect of particle acceleration on viscous drag) and added mass terms. Basset history and added mass terms were included in an elegant coupled phase formulation by Harker and Temple.[15] Coupled phase effects were also treated by Evans and Attenborough[16,17] in an extension to the work of Harker and Temple[15] to incorporate thermal conduction. Nonstationary effects, which become important at low particle Reynolds numbers, are further investigated by Gumerov *et al.*[18] Chow[19] considered the attenuation of sound in dilute emulsions and suspensions with viscous dissipation and thermal conduction and additionally included the effect of surface tension, which was shown to be important in the presence of bubbles (and negligible for systems with solid particles or liquid droplets).

Temkin and Dobbins,[20] in their classical work, theoretically considered particle attenuation and dispersion of sound in a manner which illustrates explicitly the relaxation character of the problem. For rigid particles, the linear droplet drag and heat transfer are, respectively, obtained from

$$F_p = 6\pi\mu_g(u_p' - u_g'), \tag{2a}$$

$$Q_p = 2\pi d_p k_g(T_p' - T_g'), \tag{2b}$$

which correspond to the zero droplet Reynolds number limit ($\mathrm{Re}_p \to 0$), where

$$\mathrm{Re}_p = \rho_g|u_g - u_p|d_p/\mu_g. \tag{2c}$$

In the above, u is the velocity, T the temperature, ρ the density, μ the dynamic viscosity, k the thermal conductivity, and d the particle diameter. The subscripts g and p stand, respectively, for the gas and the particle, and the primes denote fluctuations from the mean. The properties ρ_g, μ_g, k_g, and ρ_p refer to the mean values. According to Temkin and Dobbins,[20] Stokes linear drag law can be justified for $0 \leqslant \omega\tau_d \approx 1$, provided that $\rho_g/\rho_p \ll 1$ and $(\omega d_p^2/8\nu_g)^{1/2} \ll 1$.

Recent data[21] on sound attenuation in supersonic air jets containing suspended water droplets reveal that the linear absorption coefficient displays a spectral peak. The particulate relaxation models for the sound attenuation are all based on Stokes drag (linear drag law) and pure conduction limit (linear heat transfer) and do not account or explain this attenuation behavior. This article, primarily based on Ref. 22, attempts to investigate this attenuation behavior by considering nonlinear drag and heat transfer laws applicable to relatively large-sized droplets.

2. Analysis for sound attentuation

The present analysis extends the work of Temkin and Dobbins[20] for dilute suspensions to accommodate the nonlinear drag and heat transfer laws, which become important at high particle Reynolds numbers Re_p and at high frequencies. Only sound attenuation is considered here, with sound dispersion excluded from consideration.

2.1 Nonlinear particle drag and heat transfer

The theory of Temkin and Dobbins[20] for dilute solutions is applicable at *low mass concentrations* $C_m = n_0 m_p/\rho_g \ll 1$, where n_0 refers to the mean number of particles per unit volume of mixture and m_p the mass of one particle. Without any loss of generality the attenuation of sound for large particle Reynolds numbers with nonlinear particle relaxation may be expressed as follows:

$$\bar{\alpha} = \left(\frac{c_0 \alpha}{C_m \omega}\right) = \frac{\omega \tau_{d1}}{1+\omega^2 \tau_{d1}^2} + (\gamma - 1)\left(\frac{c_{pp}}{c_{pg}}\right)\frac{\omega \tau_{t1}}{1+\omega^2 \tau_{t1}^2}. \tag{3a}$$

In the above the quantity $\bar{\alpha}$ refers to the attenuation per unit frequency per unit mass fraction, α the linear absorption coefficient or the amplitude attenuation coefficient (imaginary part of the wavenumber), c_p the specific heat, c_0 the speed of sound in the gas phase, and γ the isentropic exponent (specific-heat ratio). Note that $\alpha = \alpha_i/2$.

The relaxation times τ_{d1} and τ_{t1} correspond to those under nonlinear drag conditions (generally representative of large-sized particles). They are related to the relaxation times τ_d and τ_t by the relations

$$\tau_{d1} = \tau_d \psi_1(\mathrm{Re}_p), \quad \tau_{t1} = \tau_t \psi_2(\mathrm{Re}_p, \mathrm{Pr}), \tag{3b}$$

where

$$\psi_1(\mathrm{Re}_p) = C_{D1}/C_D, \quad \psi_2(\mathrm{Re}_p, \mathrm{Pr}) = \mathrm{Nu}_1/\mathrm{Nu}, \tag{3c}$$

with C_{D1} standing for the nonlinear drag coefficient and Nu_1 for the nonlinear heat transfer. In the above,

$$\tau_d = d_p^2 \rho_p/(18\mu_g), \tag{4a}$$

and

$$\tau_t = \frac{m_p c_{pp}}{2\pi d_p k_g} = \frac{\mathrm{Pr} c_{pp} d_p^2 \rho_p}{12 \mu_g c_{pg}} = \left(\frac{3}{2}\right)\left(\frac{c_{pp}}{c_{pg}}\right)\mathrm{Pr}\,\tau_d, \tag{4b}$$

where $\mathrm{Pr} = c_{pg}\mu_g/k_g$ stands for the Prandtl number of gas. Physically the quantity τ_{d1} is a measure of the inability of the particles to follow (respond to) the fluctuations in the fluid motion.[3] Likewise, the quantity τ_{t1} is a measure of the thermal response of the particles to follow the fluctuations in the temperature of the fluid.

The drag coefficient and the Nusselt number in Eq. (3c) are defined by

$$C_D = 2F_D \Big/ \left(\rho_g \frac{\pi}{4} d_p^2 u_g^2\right), \quad \mathrm{Nu} = h_g d_p/k_g, \tag{5}$$

where F_D is the particle drag force and h_g the gas-droplet convective heat transfer coefficient. A good approximation (curve fit) for the drag coefficient C_{D1} is recommended as[23]

$$C_{D1} = \frac{24}{\mathrm{Re}_p} + \frac{6}{1+\sqrt{\mathrm{Re}_p}} + 0.4. \tag{6}$$

The first term on the right hand side of Eq. (6) is the Stokesian drag coefficient defined by

$$C_D = 24/\mathrm{Re}_p, \tag{7}$$

which is valid for particle Reynolds numbers less than about 1. Equations (5) and (6) yield an expression for ψ_1 as

$$\psi_1(\mathrm{Re}_p) = 1 + \frac{\mathrm{Re}_p}{24}\left(\frac{6}{1+\sqrt{\mathrm{Re}_p}} + 0.4\right). \tag{8}$$

With regard to the particle heat transfer, a good correlation for heat transfer (by conduction and convection) is expressed by the well-known Ranz–Marshall correlation[24]

$$\mathrm{Nu}_1 = 2 + 0.6\mathrm{Re}_p^{0.5}\mathrm{Pr}^{0.33}. \tag{9}$$

In the pure conduction limit, we have Nu=2 so that the function ψ_2 in Eq. (3c) becomes

$$\psi_2 = \mathrm{Nu}_1/\mathrm{Nu} = 1 + 0.3\mathrm{Re}_p^{0.5}\mathrm{Pr}^{0.33}. \tag{10}$$

It now remains to find a relation for the droplet Reynolds number Re_p in terms of the density ratio and the particle relaxation time.

2.2 Droplet Reynolds number

The determination of particle Reynolds number required in the evaluation of the functions ψ_1 and ψ_2 in Eqs. (8) and (10), respectively, is exceedingly complex. There exists relatively little information on the dependence of particle Reynolds number on the particle characteristics in two-phase flows. A study of turbulent diffusion of droplets in a gaseous medium[25,26] indicates a *plausible* relationship of the form

$$\mathrm{Re}_p = f(\rho_g/\rho_p, \omega\tau_d). \tag{11}$$

Since Eq. (3a) is applicable to large particle to fluid density ratios, we postulate here that the particle Reynolds number depends only on the particle relaxation time and is independent of the particle to fluid density ratio, that is,

$$\mathrm{Re}_p = f(\omega\tau_d). \tag{12}$$

The physical basis for this supposition may be justified as follows. Particles with a smaller relaxation time are able to respond better to the fluctuations in the fluid motion, which leads to reduced particle relative velocity and thus a reduced particle Reynolds number.

In the present investigation, the following power law relation is proposed such that a peak in the linear absorption coefficient (as indicated by the measurements) is realized (an exponent of 2.5 or greater for the relaxation time ensures this peak):

$$\mathrm{Re}_p = c(\omega\tau_d)^3. \tag{13}$$

The adjustable constant c is determined from a correlation of the theory with the test data. A value of $c=10$ is found to be satisfactory based on the data of Norum[21] for water droplets in a supersonic air jet.

3. Results and comparison

3.1 Effect of nonlinear particle relaxation

The effect of nonlinear particle relaxation on the absorption coefficient per unit frequency is demonstrated in Fig. 2(a) for comparison with the theory of Temkin and Dobbins[20] for dilute concentrations. The results suggest that below $\omega\tau_d \approx 0.7$ the nonlinear particle relaxation effects are seen to be unimportant. Beyond $\omega\tau_d \approx 0.7$, the nonlinear theory departs from the linear theory.

Figure 2(b) shows a comparison of the predicted linear spectral attenuation of sound under nonlinear particle relaxation. The theoretical result by Temkin and Dobbins[20] for dilute concentrations is also presented for a comparison. The theory suggests that at high frequencies

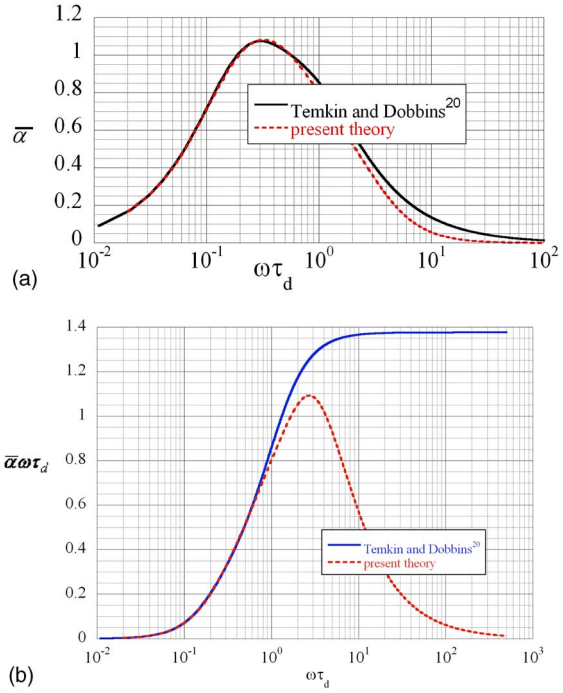


Fig. 2. (Color online) (a) Predicted absorption coefficient per unit frequency with nonlinear particle relaxation processes. (b) Predicted linear absorption coefficient with nonlinear particle relaxation processes.

the linear absorption coefficient predicted by the nonlinear theory decreases with increasing frequency in accordance with experimental trend for large particle sizes.

3.2 Comparison with experimental data

A direct comparison of the present theory with the measured spectral attenuation by Norum[21] for water droplets in a perfectly expanded (jet exit pressure equals the ambient pressure; thus shock-free) supersonic jet is displayed in Fig. 3. The data correspond to hot supersonic jet of air from a convergent-divergent nozzle operation at a jet total temperature $T_t=867$ K and a jet exit Mach number $M_j=1.45$. The jet Mach number is defined as $M_j=u_j/c_j$, where the subscript j refers to the nozzle exit conditions. The mass flow rate (maximum considered) of water to that of the jet is about 0.85. The angle θ is measured from the jet inlet axis. The jet exit Reynolds number $\mathrm{Re}_{j1}=u_j d_j/\mu_{gj}$ is about 1.3×10^6, where d_j is the jet exit diameter. At this condition,

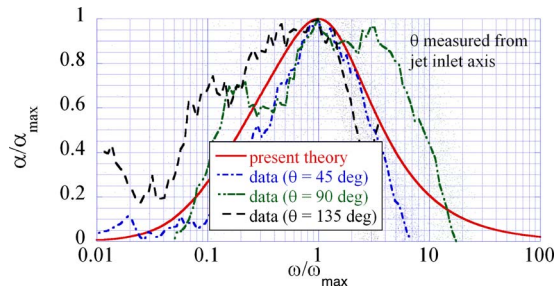


Fig. 3. (Color online) Comparison of the predictions for the linear absorption coefficient with test data of Norum (Ref. 21).

supersonic turbulent jet mixing noise dominates upstream noise radiation, and Mach wave radiation dominates the downstream noise radiation. In the data, water is injected at 45 deg. The data include spectra measured at angles of 45, 90, and 135 deg, thus highlighting directivity effects.

From the measured ΔSPL (reduction in sound pressure level) at a given frequency due to water injection, the linear absorption coefficient is deduced as

$$\alpha/\alpha_{\max} = \Delta\mathrm{SPL}/\Delta\mathrm{SPL}_{\max}, \tag{14}$$

where the subscript refers to the peak spectral reduction. The comparisons suggest that the proposed theory based on the nonlinear particle relaxation processes satisfactorily describes the measurements for the spectral attenuation of sound, indicating a spectral peak. The inclusion of directionality effect (dependence on the angle of observation) on the spectral absorption is beyond the scope of the present work.

It should be remarked, however, that the data on supersonic jet noise considered here for comparison are not directly pertinent to plane waves of sound for which the theory has been developed. Since noise from turbulent jets may be regarded as a superposition of plane waves of differing frequency, the absolute values of the absorption coefficient in the test data could be different from that expressed by the theory. In this connection, the reduction in turbulent jet mixing noise reduction by water injection was theoretically considered by the author in Ref. 27 in conjunction with the scaling laws proposed for jet noise.[28] Finally, the effect of droplet evaporation on the spectral attenuation of sound, as indicated by the theory of Marble and Candel,[29] is left out of account in the present analysis.

4. Conclusion

The theory proposed here for sound attenuation in dilute suspensions with nonlinear particle drag and heat transfer is shown to satisfactorily represent the test data for noise reduction with water droplets suspended in a supersonic jet. It is found that the nonlinear particle relaxation processes are primarily responsible for reduction in the linear absorption coefficient at high frequencies.

Acknowledgments

The author is grateful to the reviewer for helpful suggestions including several important references pertinent to suspension acoustics. Thomas Norum of NASA Langley Research Center kindly provided the spectral data for comparison. Thanks are also due to Dr. Paul Schallhorn (Chief, Environments and Launch Approval Branch) and Leonard Duncil of the Launch Services Program at NASA Kennedy Space Center for partial financial support of this work. Papers by the author from 1973 to 1982 were published with the name K. Mastanaiah.

References and links

[1]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[2]A. D. Pierce, *Acoustics: An Introduction to Physical Principles and Applications* (The Acoustical Society of America, New York, 1994).
[3]S. Temkin, *Elements of Acoustics* (The Acoustical Society of America, New York, 2001).
[4]J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. **2**.
[5]G. G. Stokes, "On the theories of internal friction of fluids in motion," Trans. Cambridge Philos. Soc. **8**, 287–305 (1845).
[6]H. von Helmholtz, "On the influence of friction in the air on sound motion," Verhandl. Naturhist. Med. Ver. Heidelberg **3**, 16–20 (1863); reprinted in Wissenschftliche Abhandlungen, Barth, Leipzig **1**, 383–387 (1882).
[7]G. Kirchhoff, "On the influence of heat conduction in a gas on sound propagation," Ann. Phys. Chem. **134**, 177–193 (1868).
[8]S. Temkin, *Suspension Acoustics* (Cambridge University Press, Cambridge, 2005).
[9]R. E. Challis, M. J. W. Povey, M. L. Mather, and A. K. Holmes, "Ultrasound techniques for characterizing colloidal dispersions," Rep. Prog. Phys. **68**, 1541–1637 (2005).
[10]C. J. T. Sewell, "On the extinction of sound in a viscous atmosphere by small obstacles of cylindrical and spherical form," Philos. Trans. R. Soc. London, Ser. A **210**, 239–270 (1910).
[11]P. S. Epstein, "On the absorption of sound waves in suspensions and emulsions," *Theodore von Karman*

Anniversary Volume (California Institute of Technology, Pasadena, California, 1941), pp. 162–188.
[12]P. S. Epstein and R. R. Carhart, "The absorption of sound in suspensions and emulsions, I. Water fog in air," J. Acoust. Soc. Am. **25**, 553–565 (1953).
[13]J. R. Allegra and S. A. Hawley, "Attenuation of sound in suspensions and emulsions: Theory and experiments," J. Acoust. Soc. Am. **51**, 1545–1564 (1972).
[14]S. L. Soo, "Effect of transport properties on attenuation and dispersion in aerosols," J. Acoust. Soc. Am. **32**, 943–946 (1960).
[15]A. H. Harker and J. A. G. Temple, "Velocity and attenuation of ultrasound in suspensions of particles in fluid," J. Phys. D: Appl. Phys. **21**, 1576–1588 (1988).
[16]J. M. Evans and K. Attenborough, "Coupled phase theory for sound propagation in emulsions," J. Acoust. Soc. Am. **102**, 278–282 (1997).
[17]J. M. Evans and K. Attenborough, "Sound propagation in concentrated emulsions: Comparison of coupled phase model and core-shell model," J. Acoust. Soc. Am. **112**, 1911–1917 (2002).
[18]N. A. Gumerov, A. I. Ivandaev, and R. I. Nigmatulin, "Sound waves in a monodisperse gas-particle or vapor-droplet mixtures," J. Fluid Mech. **193**, 53–74 (1988).
[19]J. C. F. Chow, "The attenuation of acoustic waves in dilute emulsions and suspensions," J. Acoust. Soc. Am. **36**, 2395–2401 (1964).
[20]S. Temkin and R. A. Dobbins, "Attenuation and dispersion of sound by particulate relaxation processes," J. Acoust. Soc. Am. **40**, 317–324 (1966).
[21]T. D. Norum, "Reductions in multi-component jet noise by water injection," *Tenth AIAA/CEAS Aeroacoustics Conference, Manchester*, Great Britain, May 2004, paper no. AIAA-2004-2976.
[22]M. Kandula and M. Lonergan, "Spectral attenuation of sound in dilute suspensions with nonlinear particle relaxation," *Acoustics'08 Paris Conference*, Paris, France, 29 June–4 July 2008.
[23]F. M. White, *Viscous Fluid Flow*, 2nd ed. (McGraw-Hill, New York, 1991).
[24]W. E. Ranz and W. P. Marshall, "Evaporation from drops," Chem. Eng. Prog. **48**, 141–146, 173–180 (1952).
[25]K. Mastanaiah, "Experimental and theoretical investigation of droplet deposition and heat transfer in two phase turbulent flow through a vertical tube," Ph.D. thesis, University of Illinois, 1980.
[26]E. N. Ganic and K. Mastanaiah, "Investigation of droplet deposition from a turbulent gas stream," Int. J. Multiphase Flow **7**, 401–411 (1981).
[27]M. Kandula, "Prediction of turbulent jet mixing noise reduction by water injection," AIAA J. (2008).
[28]M. Kandula, "On the scaling laws and similarity spectra for jet noise in subsonic and supersonic flow," Int. J. Acoust. Vib. **13**, 3–16 (2008).
[29]F. E. Marble and S. M. Candel, "Acoustic attenuation in fans and ducts by vaporization of liquid droplets," AIAA J. **13**, 634–639 (1975).

Parkinson's disease and the effect of lexical factors on vowel articulation

Peter J. Watson and Benjamin Munson
Department of Speech-Language-Hearing Sciences, University of Minnesota-Twin Cities, 164 Pillsbury Drive SE, Minneapolis, Minnesota 55455
pjwatson@umn.edu, munso005@umn.edu

Abstract: Lexical factors (i.e., word frequency and phonological neighborhood density) influence speech perception and production. It is unknown if these factors are affected by Parkinson's disease (PD). Ten men with PD and ten healthy men read CVC words (varying orthogonally for word frequency and density) aloud while audio recorded. Acoustic analysis was performed on duration and Bark-scaled F1-F2 values of the vowels contained in the words. Vowel space was larger for low-frequency words from dense neighborhoods than from sparse ones for both groups. However, the participants with PD did not show an effect of density on dispersion for high-frequency words.

© 2008 Acoustical Society of America
PACS numbers: 43.70.Dn [AL]
Date Received: June 6, 2008 **Date Accepted:** August 13, 2008

1. Introduction

The size of the F1-by-F2 vowel space is an important determinant of speech intelligibility. Talkers with smaller-sized vowel spaces are less intelligible than those with larger vowel spaces.[1,2] The size of the F1-by-F2 vowel space is reduced in persons with Parkinson's disease (PD) relative to healthy adults.[3,4] Consequently, there is a smaller acoustic distance between vowels. This presumably makes it more difficult for listeners to discriminate differences among vowels. Vowel space is particularly sensitive in predicting speech intelligibility in persons with PD,[4] and knowledge of factors that contribute to vowel space would be important to understand.

In neurologically healthy talkers, within-talker vowel production is not only affected by task-related variables (e.g., articulation rate and presence of background noise) but also by lexical factors such as phonological neighborhood density and word frequency. Neighborhood density is defined as the number of words that differ by one phoneme (e.g., *kin's* neighbors include *in, skin, can*, and *kiss*). Infrequently used words from dense neighborhoods are more difficult to identify than high-frequency low-density words.[5] In both younger and older adults, less frequently used words from dense neighborhoods are produced with more expanded vowel space than frequently used words from sparse neighborhoods.[6–9] Wright[6] maintained that one possible reason for this expansion of vowel space is that talkers are accommodating listeners by making vowels from these lexically difficult words more distinct, thus partially countering the intelligibility decrement that the "hard" words offer.

There is a reason to suspect an interaction of PD with lexical factors. PD is caused by a loss of dopaminergic neurons within the basal ganglia. Ullmann[10] proposed that the basal ganglia are involved in the selection and organization of phonemes for word production, and the basal ganglia have been implicated in phonological processing in neuroimaging and neuroelectrophysiological studies of awake-healthy humans.[11–13]

The identification of acoustic variables predictive of poor intelligibility in persons with PD has not yielded satisfactory results.[4,14] The purpose of this preliminary investigation was to examine the effect of word frequency and neighborhood density on vowel production in persons with PD. A finding that persons with PD do not modify their vowel production in response to word frequency and neighborhood density would motivate future studies of linguistic influences on speech production and its relations to intelligibility.

Table 1. Stimulus words used.

Vowel	High frequency/ high density	High frequency/ low density	Low frequency/ high density	Low frequency/ low density
ɑ	got	dock	dot	mop
ɑ	pot	top	cot	cop
æ	sad	sang	fad	sag
ɛ	get	death	dept	deaf
eɪ	game	gain	dame	babe
i	beat	beach	beak	leach
ou	note	wrote	moat	rope
u	moon	room	womb	tune

2. Methods

Ten men diagnosed with idiopathic PD participated and were of an average age of 73 years (SD=9.5 years). A group of ten healthy men, with an average age of 72 years (SD =7.7 years), served as healthy controls (HCs). Data from five of these HC participants had been analyzed and reported previously.[9] Of the participants with PD, four had Hoehn-Yahr[15] scores of 3 and six had scores of 4. All participants with PD were judged by the first author (a certified speech-language pathologist) to have dysarthria. Participants with PD scored 26 or higher on the Mini-Mental-Scale-Examination. Audio recording of persons with PD was done when they reported they were "on" state, 1–2 h after taking their medications. All participants were native American-English talkers.

In order to shorten the elicitation session to minimize possible fatigue, the stimuli were 32 CVC words with monophthongal vowels (/i/, /a/, /eɪ/, /ɛ/, /æ/, /oʊ/, and /u/) from a subset of the 80 words used by Munson and Solomon.[7] Four lists were constructed with high- and low-frequency words coming from phonological neighborhoods of high and low densities. The lists were constructed by conducting a median split for the raw frequency and frequency-weighted neighborhood density values for the CVC words in the Hoosier-Mental-Lexicon[16] and constructing phonemically matched quadruplets of words containing the same vowel.[17] The average occurrence of the high-frequency words was 182, while the average for the low-frequency words was 7. This difference was significant in an ANOVA ($F[1,28]=11.4$, $p=0.002$). The average frequency-weighted neighborhood density was 188.4 for the high-density words and 32.6 for the low-density words. This was significant, $F[1,28]=49.5$, $p<0.001$. Though the high- and low-density lists were not designed to contrast in raw neighborhood density, the two lists did differ marginally in this variable, $F[1,28]=3.2$, $p=0.084$. The high-density words had on average 3 more neighbors than the low-density ones. See Table 1 for the words used.

Participants read the words aloud from cards in quasiransom order. Each word was read three times, for a total of 96 productions per participant. Each participant was seated and wore a head-mounted condenser microphone (AKG, 420) interfaced with a digital-CD recorder (Marantz, DRR300). The acoustic data were recorded at 44.1 kHz sampling frequency with 16 bit quantization.

Vowel segments were determined from both the waveform and spectrogram displays with linear predictive coding superimposed over the spectrogram display using PRAAT software.[18] From the marked segments, vowel duration and Bark-scaled F1 and F2 values[18] were extracted. For each vowel, mean vowel duration, mean F1, mean F2, and mean vowel-space dispersion were calculated. The mean F1 and F2 values collapsed for all vowels provided an average-centroid value within the vowel space. The calculation of vowel dispersion was determined by taking the average Euclidean distance of the individual vowels from the center of the vowel space, defined operationally as the average F1 and F2 for the entire set of vowels. A

Table 2. Means and standard error of the mean (in parenthesis), in milliseconds, for the HC and the PD groups.

	High frequency/ high density	High frquency/ low density	Low frequency/ high density	Low frequency/ low density
HC	219.62 (9.29)	210.69 (8.23)	211.09 (10.97)	288.05 (7.51)
PD	246.43 (23.53)	240.18 (22.62)	233.57 (16.40)	230.82 (20.46)

three-factor mixed-model ANOVA, with participant group as the between factor and word frequency and density as the within factors, was used to test vowel duration and dispersion.

3. Results

Vowel duration for the HC and PD groups are shown in Table 2. On average, vowel duration was longer for the PD group, but the ANOVA showed no main effect for group due most likely to the large variability within the PD group. There was a significant frequency by group interaction [$F(1,18)=6.288$, $p=.022$, $\eta^2=0.259$], showing that vowels in high-frequency words were longer for the PD group only.

Vowel dispersion for high- and low-frequency and high- and low-density words are shown in Fig. 1. As can be seen, average dispersion was larger in the HC group for all words. The ANOVA showed a significant main effect for group [$F[1,18]=11.93$, $p=.003$, $\eta^2=0.40$]. There were also a significant effect of density ($F[1,18]=13.95$, $p=0.002$, $\eta^2=0.44$), a significant frequency by density interaction ($F[1,18]=12.68$, $p=0.002$, $\eta^2=0.41$), and a significant group by frequency by density interaction [$F(1,18)=5.843$, $p=0.026$, $\eta^2=0.25$]. Follow-up testing with Bonferroni-corrected independent sample t-tests revealed significant group differences in dispersion for all four stimulus types. The three-way interaction can be seen by comparing the bar heights in Fig. 1. Bonferroni-corrected t-tests showed that the HC talkers produced more-dispersed vowels in high-density words than low-density words regardless of whether they were high or low in frequency [$t(9)=3.2$ and 2.9 and $p=0.011$ and 0.018, respec-

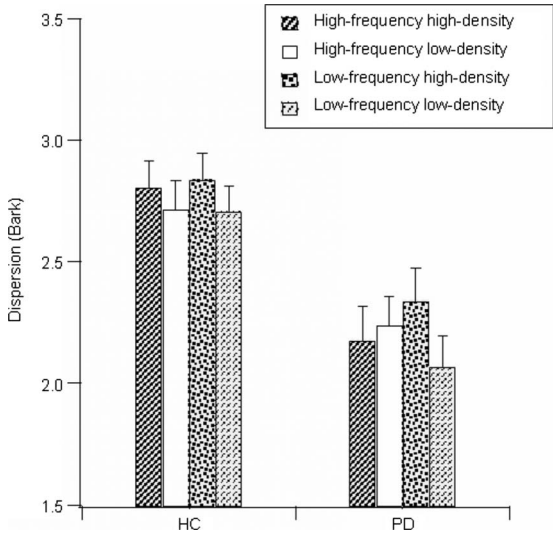


Fig. 1. Mean F1/F2 vowel dispersion and standard error of the mean (error bars) in Bark, for high-frequency high-density, high-frequency low-density, low-frequency high-density, and low-frequency low-density words for the HC and PD groups.

tively]. There were no effects of word frequency. For the PD talkers, there was an effect of density for the low-frequency words only [$t(9)=3.0$, $p=0.015$]. Moreover, the high-frequency low-density words were less dispersed than the low-frequency low-density words [$t(9)=3.24$, $p=0.01$].

Finally, a set of regression analyses examined whether the influence of group, frequency, and density on vowel-space dispersion was an artifact of vowel duration, as previous research has shown these variables to be correlated.[19] This was accomplished by conducting a regression predicting average dispersion from average duration. Eighty data points were used in this regression, representing the four data points for all 20 participants. The regression was significant, but with only 6% of the variance in dispersion being attributable to duration. Furthermore, this relation was in the opposite-than-predicted direction, owing largely to the fact that the individuals with PD produced both longer durations and less-expanded vowel spaces.

4. Discussion

These findings suggest that the effect of PD on speech produces a more complex interaction of vowel production than has previously been described. Vowel-space dispersion was contracted in the PD group for this current study, which is consistent with numerous other studies of the speech of individuals with PD (e.g., Refs. 3, 4, and 14). Vowel-space dispersion was not an artifact of duration-dependent undershoot, as demonstrated by the inverse relation between duration and dispersion in the PD group.

Even though vowel dispersion in the PD group was contracted for all words, they did show some of the same lexical effects that had been demonstrated previously in the speech of individuals without speech, language, or hearing disorders. The PD group, for example, showed the expected effect of density on vowel-space dispersion for low-frequency words, though not for high-frequency words. That the effect of density would be limited to low-frequency words is consistent with Wright's[6] initial speculation that these effects derived from tacit attempts to enhance perceptibility of words, as low-frequency words are predicted overall to be less intelligible than high-frequency ones. It is also consistent with the pattern found for a larger group of healthy older adults analyzed by Watson and Munson.[9] The reason that this pattern was not observed in high-frequency words is less clear, but we offer three possible explanations. First, as suggested by Munson,[8] word-frequency effects appear more vulnerable to task-related influences than those of density and would be less consistent across studies and corpora. Second, a recent study[20] showed that persons with PD had impaired recall for high-frequency but not for low-frequency words when compared to healthy age- and gender-matched controls. It is possible that persons with PD are less sensitive to density effects when producing high-frequency words. Third, although all participants with PD were identified as having dysarthria, no determination of speech impairment severity was made, which may have had an effect on density discrimination in high-frequency words. However, further study is required to determine the reason for this effect.

Most importantly, however, the fact that the talkers with PD did not show the expected patterns of density on dispersion for the high-frequency words may compromise their intelligibility. This finding suggests to us two dual sets of follow-up studies. The first of these should examine whether the failure to modulate the dispersion of high- and low-density high-frequency words is maintained in the connected speech of individuals with PD and, if so, if this failure contributes to their decreased speech intelligibility. Second, studies should examine whether this failure can be linked to a difficulty in the formulation of speech related to problems of phonological processing.

References and links

[1]A. Bradlow, G. Torretta, and D. Pisoni, "Intelligibility of normal speech I: Global and fine-grained acoustic phonetic talker characteristics," Speech Commun. **20**, 255–272 (1996).
[2]A. Neel, "Vowel space characteristics and vowel identification accuracy," J. Speech Lang. Hear. Res. **51**, 574–585 (2008).
[3]J. Liss, S. Spitzer, J. Caviness, C. Adler, and B. Edwards, "Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech," J. Acoust. Soc. Am. **104**, 2457–2466 (1988).

[4]G. Weismer, J. Jeng, J. Laures, R. Kent, and J. Kent, "Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders," Folia Phoniatr Logop **53**, 1–18 (2001).
[5]P. Luce and D. Pisoni, "Recognizing spoken words: The neighborhood activation model," Ear Hear. **19**, 1–36 (1998).
[6]R. Wright, "Factors of lexical competition in vowel articulation," in *Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge, 2004), pp. 26–50.
[7]B. Munson and N. P. Solomon, "The effect of phonological neighborhood density on vowel articulation," J. Speech Lang. Hear. Res. **47**, 1048–1058 (2004).
[8]B. Munson, "Lexical access, lexical representation, and vowel production," in *Laboratory Phonology IX*, edited by J. Cole and J. S. Haulde (Mouton de Gruyter, New York, 2007), pp. 201–228.
[9]P. J. Watson and B. Munson, "A comparison of vowel acoustics between older and younger adults," *Proceedings of the 16th International Congress on Phonetic Sciences* (University of Saarland, Saarbrucken, Germany, 2007).
[10]M. Ullman, "A neurocognitive perspective on language: The declarative/procedural model," Nat. Neurosci. **2**, 717–726 (2001).
[11]J. Booth, L. Wood, D. Lu, J. Houk, and T. Bitan, "The role of the basal ganglia and cerebellum in language processing," Brain Res. **1133**, 136–144 (2007).
[12]M. Tettamanti, A. Moro, C. Messa, R. Moresco, G. Rizzo, A. Carpinelli, M. Matarrese, F. Fazio, and D. Perani, "Basal ganglia and language: phonology modulates dopaminergic release," NeuroReport **16**, 397–401 (2005).
[13]Y. Abdullaev and K. Melnichuk, "Cognitive operations in the human caudate nucleus," Neurosci. Lett. **234**, 151–155 (1997).
[14]K. Tjaden and G. E. Wilding, "Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings," J. Speech Lang. Hear. Res. **47**, 766–783 (2004).
[15]M. Hoehn and M. Yahr, "Parkinsonism: Onset, progression, and mortality," Neurology **17**, 427–442 (1967).
[16]D. Pisoni, H. Nusbaum, P. Luce, and L. Slowiaczek, "Speech perception, word recognition, and the structure of the lexicon," Speech Commun. **4**, 75–95 (1985).
[17]As previously reported in Eq. (8), the neighborhood densities of the stimuli in Eq. (7) were described incorrectly. For additional details on the correct description of these stimuli, see Eq. (8).
[18]P. Boersma and W. Weenik, PRAAT, a system for doing phonetics by computer (Version 4.4.30) (Institute of Phonetic Sciences, Amsterdam, 2005).
[19]S. Moon and B. Lindblom, "Interactions between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55 (1994).
[20]P. Davidson, D. Anaki, J. Satin-Cyr, T. Chow, and M. Moscovitch, "Exploring recognition memory deficit in Parkinson's disease: Estimates of recollection versus familiarity," Brain **129**, 1768–1779 (2006).

High-frequency dispersion from viscous drag at the grain-grain contact in water-saturated sand

Nicholas P. Chotiros and Marcia J. Isakson
Applied Research Laboratories, The University of Texas at Austin, Texas 78713-8029
chotiros@arlut.utexas.edu, misakson@arlut.utexas.edu

Abstract: Shear viscous drag within the thin fluid film at the contact between grains in water-saturated sand is an important loss mechanism for high-frequency sound in the Biot–Stoll plus contact squirt flow and shear viscous drag (BICSQS) model [J. Acoust. Soc. Am. **116**, 2011–2022 (2004)]. Couette flow was assumed for the shear drag but it breaks down when inertial effects within the film become significant. Using Biot's method, a correction is derived for the shear drag and inserted into the BICSQS model. The result is a prediction of negative sound speed dispersion, consistent with dynamic theories of fluid-filled poroelastic bodies.

© 2008 Acoustical Society of America
PACS numbers: 43.20.Jr, 43.30.Ma [AN]
Date Received: May 29, 2008 **Date Accepted:** August 13, 2008

The Biot model[1] has been successfully applied to fused glass beads, but it is unable to predict the attenuation and sound speed dispersion measured in unconsolidated water-saturated sand over a broad band of frequencies.[2] The difficulty is in the grain contact physics, which is more complicated than that of a fused link. The measured data were more consistent with the Biot–Stoll plus contact squirt flow and shear viscous drag (BICSQS) model,[2] which represents the grain contact physics in terms of a normal and a tangential contact stiffness. The normal contact stiffness is modeled as a simple spring modified by squirt flow. The tangential contact stiffness is modeled as a simple spring with shear drag, in which Couette flow is assumed. Inertial effects will cause deviations from Couette flow at high frequencies. This paper addresses the high-frequency correction for the tangential contact stiffness.

For a fluid film of thickness $2a_1$ between two parallel, solid, planar surfaces moving in parallel but opposite directions, at constant velocities A and $-A$, respectively, the shear stress τ at each surface is given by

$$\tau = A\mu/a_1, \tag{1}$$

where μ is the viscosity of fluid. The fluid velocity profile is simple; it changes linearly from $-A$ to A with distance across the fluid film. For a slow sinusoidal motion, this expression is still approximately correct, but at high frequencies it will break down due to the inertial effects within the fluid. Therefore, a correction function $F_{1s}(\alpha)$ needs to be applied to the above equation, in which α is the angular frequency; thus,

$$\tau = F_{1s}(\alpha)A\mu/a_1. \tag{2}$$

The nomenclature is similar to that of Biot because the derivation is inspired by his derivation of the drag force of an oscillatory flow in a two-dimensional duct, specifically Eqs. (2.1)–(2.17) of Ref. 1. Considering two-dimensional motion, in which the x-direction is parallel to the plane boundaries, and the y-axis is perpendicular to them and where the boundaries are at $y = \pm a_1$, the equation of the fluid velocity u_{1s} in the x-direction is

 © 2008 Acoustical Society of America

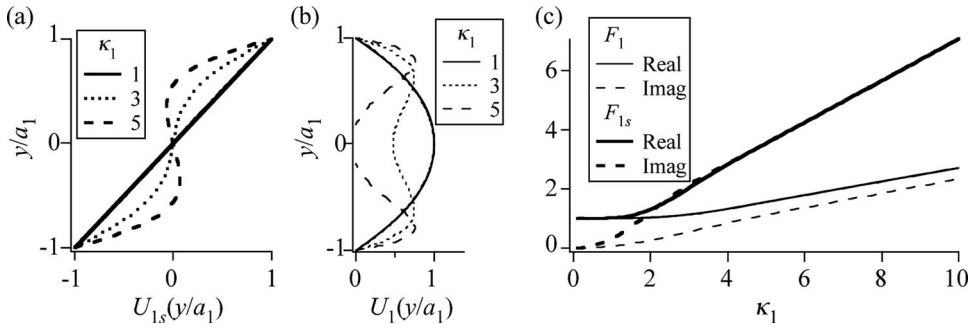


Fig. 1. Real part of (a) flow profiles in a sinusoidal shear film U_{1s}, (b) flow profiles U_1 from Ref. 1 at selected values of κ_1, and (c) complex viscous drag correction functions F_1 and F_{1s}.

$$\rho_f \frac{\partial u_{1s}}{\partial t} = \mu \frac{\partial^2 u_{1s}}{\partial y^2}, \tag{3}$$

where ρ_f is fluid density. Let the fluid velocity be defined as a product of a sinusoidal oscillation and a profile function $U_{1s}(y)$ as follows:

$$u_{1s} = U_{1s}(y)e^{i\alpha t}, \quad U_{1s}(\pm a_1) = \pm A, \tag{4}$$

where α is the angular frequency, t is time, and A is the amplitude of the sinusoidal velocity at the boundaries. Substituting into Eq. (3), the result is

$$i\alpha U_{1s}(y) = \nu \frac{\partial^2 U_{1s}(y)}{\partial y^2}, \tag{5}$$

where $\nu = \mu/\rho_f$ is the kinematic viscosity. The general solution is of the form

$$U_{1s}(y) = B \sinh(y\sqrt{i\alpha/\nu}) + C \cosh(y\sqrt{i\alpha/\nu}), \tag{6}$$

where B and C are constants. For the case of flow drag between parallel plane boundaries, the expected solution is an even function, and B would be zero, as in Ref. 1. In the case of shear drag, the expected solution must be an odd function, and C would be zero. An expression for B may be found from the boundary conditions. Following Biot's method, the variable α is replaced with κ_1, according to the definition

$$\kappa_1 = a_1\sqrt{\alpha/\nu}. \tag{7}$$

The resulting solution for the fluid velocity profile is

$$U_{1s}(y) = A\frac{\sinh\left(\frac{y}{a_1}\kappa_1\sqrt{i}\right)}{\sinh(\kappa_1\sqrt{i})}. \tag{8}$$

Normalized examples of the real part of U_{1s} at three values of κ_1 are shown in Fig. 1(a) as an illustration. At low frequencies, it is almost linear, but at high frequencies inertial effects come into play. The imaginary part, not shown, tends to zero at low frequencies. For comparison, the corresponding profiles of U_1 for sinusoidal flow between parallel boundaries from Ref. 1 are shown in Fig. 1(b). Omitting the frequency dependent term $e^{i\alpha t}$, the shear stress at the boundaries is given by

$$\tau = \mu \frac{\partial U_{1s}(y)}{\partial y}\bigg|_{y=\pm a_1} = \mu \frac{A}{a_1} \frac{\kappa_1 \sqrt{i}\cosh(\kappa_1 \sqrt{i})}{\sinh(\kappa_1 \sqrt{i})}. \tag{9}$$

Eliminating τ from Eqs. (2) and (9), an expression for the high-frequency correction is obtained as follows:

$$F_{1s}(\kappa_1) = \frac{\kappa_1 \sqrt{i}\cosh(\kappa_1 \sqrt{i})}{\sinh(\kappa_1 \sqrt{i})}. \tag{10}$$

The behavior of the function is similar to that of Biot's correction $F_1(\kappa_1)$ for the flow drag between parallel boundaries; both functions start at (1,0) at low frequencies and their real and imaginary parts increase monotonically with frequency but at different rates. Plots of both functions are shown in Fig. 1(c).

This result may be used to improve the BICSQS model[2] at high frequencies. Specifically, the expression for the frequency dependent tangential contact stiffness S_t between a pair of solid grains was originally given in Eq. (12) of Ref. 2 as

$$S_t = \frac{A_S}{A_\theta} = g_c\left(1 - i\frac{\omega}{\omega_\mu}\right), \tag{11}$$

where A_θ is the amplitude of a sinusoidal tangential strain, A_S is the corresponding tangential stress, g_c is the stiffness due to solid contact, and the imaginary term is the shear drag due to viscosity within the fluid film that is trapped between the grain surfaces. Assuming that the fluid film at the grain-grain contact may be approximated as a thin fluid film between a pair of plane boundaries, the correction term may be simply applied to the shear drag term. The new expression is

$$S_t = \frac{A_S}{A_\theta} = g_c\left(1 - iF_{1s}(\kappa_1)\frac{\omega}{\omega_\mu}\right). \tag{12}$$

Following the same logic as in Ref. 2, the frame shear modulus becomes

$$\mu = G_c\left(1 - iF_{1s}(\kappa_1)\frac{\omega}{\omega_\mu}\right). \tag{13}$$

The effect of this modification will be illustrated by way of an example, based on the BICSQS model of the sediment at the Sediment Acoustics Experiment of 1999 (SAX99), and later revisited in 2004 (SAX04). The model parameters are provided in the last column of Table IV in Ref. 2. One additional parameter is needed, i.e., the value of κ_1. As defined in Eq. (7), the gap between the solid surfaces is the main unknown. It has to be of the same order of magnitude as the grain surface roughness. If the grains may be approximated as randomly close-packed spheres, which have an ideal porosity of 0.36, a rough estimate may be obtained from the difference between this and the measured porosity. The measured[3] porosity is 0.37 and the average grain diameter is about 0.4 mm. From simple geometry, the average half gap width, a_1, is calculated to be $((1-0.36)/(1-0.37))^{1/3}-1$ times the average grain radius, which comes to 0.001 mm. This should be regarded as a very rough estimate since the grains are not spherical, as shown by the scanning electron microscope images in Ref. 3.

By replacing Eq. (12) in Ref. 2 with Eq. (13) from above, the high-frequency correction was introduced into the BICSQS model. Using the equations and procedures described in Ref. 2 along with the estimated value of a_1 above, the modeled sound and shear wave speeds and attenuations for the SAX99 experiment, with and without the high-frequency correction F_{1s}, are compared in Fig. 2. The F_{1s} correction appears to have a negligible effect on the wave attenuations. It reinforces the rapid increase in shear wave speed at high frequencies. New shear wave experimental measurements will be needed for verification, since there are no data in that frequency range for unconsolidated sand.

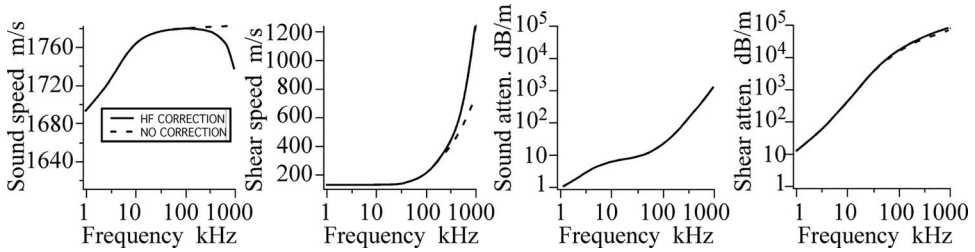


Fig. 2. Sound and shear wave speeds and attenuations computed by the BICSQS model of the SAX99 sediment, with (solid line) and without (broken line) the high-frequency correction F_{1s}.

The most noticeable effect of the F_{1s} correction is negative dispersion of the sound wave speed at high frequencies. The speed is a function of both the bulk and shear moduli; the imaginary part of F_{1s} appears to interact with 'i' in Eq. (13) to give a net reduction in the real part of the modulus for the sound wave, which results in negative dispersion. Negative dispersion has been observed in a number of experiments. It has been observed in a bone.[4] Chakraborty[5] modeled it as a nonlocal version of Biot's theory, but the internal structure of a bone is very different from that of sand. With regard to granular media, Schwartz and Plona[6] measured the sound speed and attenuation of dry and water-saturated glass and plexiglass beads between 300 kHz and 2 MHz, and matched the results to multiple scattering models. Their measurements show speed decreasing as f^{-3} and attenuation increasing as f^4; their models approximately fit the data, although they predicted a speed decrease that goes as f^{-2}. Attenuation in well-sorted, vibration-compacted, water-saturated sand between 100 kHz and 1 MHz measured by Nolle *et al.*[7] increased as $f^{1/2}$, consistent with a viscous loss model due to relative motion between pore fluid and the sand grains, similar to Biot's theory. Speed dispersion was not addressed in the experimental measurements. Sound speeds in water-saturated laboratory sands between 300 kHz and 1 MHz, measured by Guillon *et al.*[8] and by Lee *et al.*,[9] decreased as f^{-2}. The former did not show any attenuation measurements, and the attenuation measurements provided by the latter appear to be much lower than other comparable measurements.

In the BICSQS model, high-frequency sound wave attenuation is dominated by viscous shear drag in the thin fluid film at the grain-gain contact, and it predicts an attenuation that increases as f^2. When the high-frequency correction F_{1s} is included, it also predicts a speed reduction that is proportional to f^{-2}. It matches the sound speed and attenuation measurements that were made on core samples of natural sand at the SAX04 experiment by Zimmer *et al.*[10] between 60 and 400 kHz, as shown in Fig. 3. The relatively large sound speed values at 60 kHz are attributed to a systematic error resulting from the short travel distance in the sediment core relative to the acoustic wavelength at that frequency and should be ignored. The measurements show a relatively large spread that is typically observed in unconsolidated water-saturated sands. The frequency range of the sound speed measurements is not large enough, relative to the spread, to unequivocally show negative dispersion. The attenuation measurements more convincingly show the tendency toward f^2 dependence above about 100 kHz, consistent with the viscous shear drag losses in the thin fluid film at the grain-gain contact, rather than the fourth power dependence of the multiple scattering models, or the half power dependence of the Biot models with elastic frames.

In conclusion, three loss mechanisms are identified for water-saturated beads and sands at high frequencies, distinguished mainly by the frequency dependence of attenuation. (a) The viscous loss mechanism due to pore fluid flow, which gives an attenuation that increases as $f^{1/2}$ as in the Biot theory, was found in well-sorted laboratory sands that were compacted by vibration, between 100 kHz and 1 MHz, as in Ref. 7. A more recent dynamic poroelastic model by Lopatnikov and Cheng,[11] which improves upon the Biot model by taking into account porosity gradients and dynamic effects and predicts negative dispersion, may be included in the same category. (b) Multiple scattering, which appears to dominate in dry and water-saturated glass

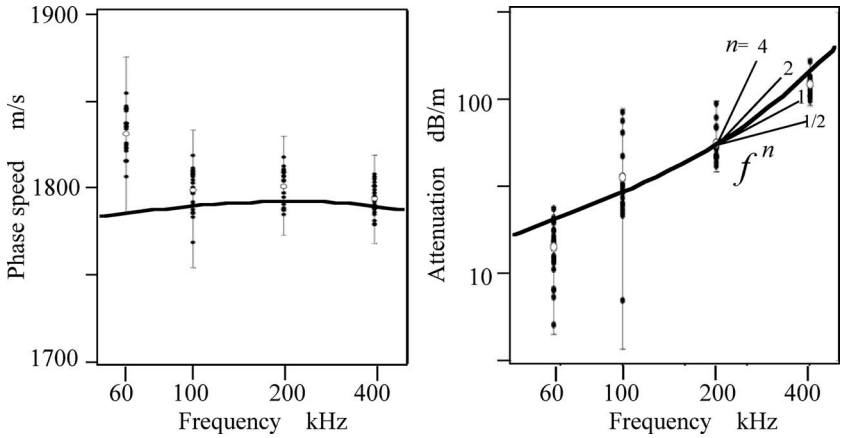


Fig. 3. BICSQS model plus the high-frequency correction F_{1s} (solid line) compared to the measurements of phase speed and attenuation in core samples from SAX04 by Zimmer *et al.* (Ref. 10) (open circles are average values and error bars represent two standard deviations).

and plexiglass beads, as in Ref. 6, predicts attenuation that increases as f^4. (c) The loss mechanism due to viscous shear drag in the fluid film at the grain-grain contact, as implemented in the BICSQS model, predicts a f^2 dependence. In this paper, a high-frequency correction is added, which does not change the frequency dependence of the attenuation, but changes the frequency dependence of the phase speed by giving it a negative dispersion. The combination matches the measurements in natural sand by Zimmer *et al.*[10] from SAX04. Although the measurements do not unequivocally show negative dispersion due to the limited frequency range, they clearly show the f^2 dependence of the attenuation, which is distinct from the Biot models with elastic frames and multiple scattering models.

Acknowledgments

The authors would like to thank Michael Zimmer, Dale Bibee, and Michael Richardson of the Naval Research Laboratory (NRL) for providing the sound speed and attenuation measurements, and to acknowledge their colleagues in the SAX04 experiment, particularly chief scientist Eric Thorsos, Applied Physics Laboratory, University of Washington (APL/UW), and the scientists and students from all the participating institutions for their cooperation. Thanks to Kang Il Lee, Kangwon National University, Republic of Korea, for clarification of the attenuation measurements in Ref. 9. This work is sponsored by the Office of Naval Research, Code 321 OA, Ocean Acoustics Program, under the management of Robert Headrick.

References and links

[1]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid—II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[2]N. P. Chotiros and M. J. Isakson, "A broadband model of sandy ocean sediments: Biot-Stoll with contact squirt flow and shear drag," J. Acoust. Soc. Am. **116**, 2011–2022 (2004).
[3]A. H. Reed, K. B. Briggs, and D. L. Lavoie, "Porometric properties of siliciclastic marine sand: A comparison of traditional laboratory measurements with image analysis and effective medium modeling," IEEE J. Ocean. Eng. **27**, 581–592 (2002).
[4]K. A. Wear, "Group velocity, phase velocity, and dispersion in human calcaneus in vivo," J. Acoust. Soc. Am. **121**, 2431–2437 (2007).
[5]A. Chakraborty, "Prediction of negative dispersion by a nonlocal poroelastic theory," J. Acoust. Soc. Am. **123**, 56–67 (2008).
[6]L. Schwartz and T. J. Plona, "Ultrasonic propagation in close packed disordered suspensions," J. Appl. Phys. **55**, 3971–3977 (1984).
[7]A. W. Nolle, W. A. Hoyer, J. F. Mifsud, W. R. Runyan, and M. B. Ward, "Acoustical properties of water-filled sands," J. Acoust. Soc. Am. **35**, 1394–408 (1963).
[8]L. Guillon, A. Moussatov, B. Brouard, and C. Ayrault, "Propagation of sound in sands: Measurements and modeling," *Oceans'98 IEEE/OES Conference Proceedings*, 98CH36259, Nice, France, 28 September–1

October 1998, pp. 380–384, Paper No. 98CH36259.
[9]K. I. Lee, V. F. Humphrey, B.-N. Kim, and S. W. Yoon, "Frequency dependencies of phase velocity and attenuation coefficient in a water-saturated sandy sediment from 0.3 to 1.0 MHz," J. Acoust. Soc. Am. **121**, 2553–2558 (2007).
[10]M. A. Zimmer, L. D. Bibee, and M. D. Richardson, "Measurement of the frequency dependence of the sound speed and attenuation of seafloor sands from 1 to 400 kHz," IEEE J. Ocean. Eng. (2007), submitted.
[11]S. L. Lopatnikov and A. H.-D. Cheng, "Macroscopic Lagrangian formulation of poroelasticity with porosity dynamics," J. Mech. Phys. Solids **52**, 2801–2839 (2004).

Psychometric functions for hybrid difference discrimination/increment detection tasks

Daniel Shepherd,[a] Michael J. Hautus, and Hamish Hutchinson
The Department of Psychology, University of Auckland, 10 Symonds Street, Auckland 1020, New Zealand
daniel.shepherd@auckland.ac.nz, m.hautus@auckland.ac.nz, hhut005@ec.auckland.ac.nz

Abstract: Psychometric functions collected from difference discrimination tasks typically have slopes near 1, whereas those from increment detection tasks have slopes near 2. Experiments exploring the effect of stimulus configuration on the psychometric function were undertaken. Some stimuli were configured to conform to difference discrimination and increment detection tasks, while other "hybrid" stimulus configurations had physical properties associated with both tasks. The results suggest that these hybrid configurations may have psychophysical properties that fall in between those found for difference discrimination and increment detection tasks.

© 2008 Acoustical Society of America
PACS numbers: 43.66.Fe, 43.66.Ba, 43.66.Dc [QF]
Date Received: July 11, 2008 **Date Accepted:** August 25, 2008

1. Introduction

Researchers have a range of psychophysical procedures and stimulus configurations at their disposal to assist in the estimation of how well an observer can discriminate between a stimulus A (the masker expressed in amplitude units) and that stimulus plus an increment $A+\Delta A$. These procedures are commonly referred to as discrimination tasks. For the two-alternative forced-choice (2-AFC) procedure, a number of different stimulus configurations are possible for discrimination tasks (Laming, 1986). If a sinusoidal masker A is continuous with $A+\Delta A$ in the temporal domain, then the observer is said to be participating in an increment detection task [Fig. 1(a)]. If A and $A+\Delta A$ are separated temporally [Fig. 1(b)], then the task is difference discrimination. Laming (1986) extended the difference discrimination task to include pedestal discriminations, which are difference discriminations with the addition of either a gated or continuous random noise background.

An additional configuration is also possible, constituting a hybrid between increment detection and difference discrimination [Fig. 1(c)]. Such a configuration is not a fanciful notion and has appeared in the literature. Moore *et al.*(1999) used a 3-AFC task in which the masker was a 1000 Hz sinusoid of 70 dB SPL, and duration 300 ms greater than that of the increment. The increments had durations of 5, 10, 20, 50, 100, or 200 ms, and always commenced 200 ms after the onset of the masker, which continued for 100 ms after the offset of the increment. The stimuli presented in each observation interval were of equal duration, and the observer was required to indicate which interval contained the increment. Should this task be considered increment detection or difference discrimination? The observer must choose between three separate stimuli to indicate which contained the increment. This suggests difference discrimination. However, the auditory feature that is to be detected is an increment in a masker of longer duration than the increment: an increment detection. A literature search did little to settle the issue, and a parting comment in a paper by Green *et al.* (1979) on the difference between gated (i.e., difference discrimination) and continuous (i.e., increment detection) pedestal maskers served only to further reinforce our curiosity:

"For how long must a gated masker be left on before it behaves like a continuous masker? We made several attempts to measure this interval with a 100-ms signal by varying masker

[a]Author to whom correspondence should be addressed.

 © 2008 Acoustical Society of America

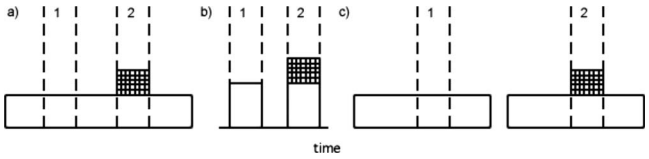


Fig. 1. Three configurations of stimuli amenable to a 2-AFC discrimination task: (a) increment detection, (b) difference discrimination, and (c) a hybrid configuration. Numbers represent observation intervals, the open rectangles the pedestal (i.e., A), and the crosshatched squares an increment (i.e., ΔA).

durations from 100 to 1600 ms. Signals were either centered in the masker, or were presented at a fixed onset asynchrony of 500 ms (McFadden, 1966). Our attempts were not entirely successful. Even at a duration of 1600 ms, the gated masker seemed somewhat more effective than the continuous masker." (p. 1055).

We conducted a simple experiment to further examine the psychophysical properties of hybrid stimuli similar to that illustrated in Fig. 1(c). Unlike Green *et al.* (1979), who varied masker duration while increment duration remained fixed, we manipulated increment duration and kept masker duration fixed. To provide an empirical test of the effect of our stimulus configurations, we employed Laming's (1986) sensory analytical model. The sensory analytical model differentiates between two classes of psychophysical tasks that yield different empirical results: the increment detection task [Fig. 1(a)] and the difference discrimination task [Fig. 1(b)]. The model predicts that the increment detection psychometric function is a normal integral with respect to the square of ΔA, while that for difference discrimination will be a normal integral with respect to ΔA.

The psychometric function for discriminating between the two tones, A and $A+\Delta A$, can be explicitly quantified by the equation

$$p(c) = \Phi\left(\left(\frac{x}{\sqrt{2}}\right)^{k}\right), \tag{1}$$

where $p(c)$ is proportionally correct in a 2-AFC procedure, $\Phi(\ )$ is the normal probability integral, x here is $\Delta A/A$,[1] and the slope exponent k is usually estimated from the data [see Laming, 1986, Eq. (4.2)]. Laming's (1986) model predicts that k will be approximately 1 for difference discriminations and approximately 2 for increment detections. When estimated from empirical data, k is usually a noninteger. The predictions of Laming's model with respect to k concur with experimental results (e.g., Irwin, 1989).

Our experiment employed trials made up of two observation intervals, one containing a 500 ms tone masker (i.e., A) and the other containing a 500 ms tone masker and an increment (i.e., $A+\Delta A$). The increment was always temporally centered in the masker. The variable manipulated in the study is increment duration, which ranged from 10 to 500 ms. For a 10 ms increment, we would expect, on the tentative assumption that the task is increment detection, the exponent to be approximately 2. For the 500 ms increment (masker and increment duration identical), the task is clearly difference discrimination, and the exponent should be about 1. By varying the increment durations between these two distinct configurations, we obtain a range of hybrid configurations. Thus what is also being tested is where increment detection begins and difference discrimination ends, the very question that interested Green *et al.* (1979).

2. Method

2.1 Observers

Three males, HH (21 years), DS (34 years), and IW (32 years), participated. Each observer underwent an audiometric assessment and none exhibited evidence of clinical hearing loss in the ear to be tested.

2.2 Stimuli and apparatus

Stimuli were 1000 Hz sinusoids generated digitally using LABVIEW 8.5. The masker A was of 500 ms duration with 50 ms ramps ($\cos^2$). The five increment durations were 10, 50, 100, 200, and 500 ms, each with 5 ms ramps ($\cos^2$). The increment was always temporally centered in the masker. The masker and increment were rendered on two channels of a 24 bit converter (National Instruments PXI-4461). The masker (Marconi, TF2162) and increment (TDT, PA5) were attenuated before being added in phase (TDT, SM5). For observation intervals containing no increment, the increment attenuator was set to its highest value. The pedestals were presented to the observer at 70 dB SPL via a monaural headset (TDH, 49P) connected to a headphone buffer (TDT, HB5). The observer was seated in a sound-attenuating chamber (Amplaid, Model E).

2.3 Procedure

Observers indicated, on a standard numeric keypad, which observation interval in the 2-AFC task contained the increment. Stimuli in a trial were separated by 200 ms. Warning, observation, and feedback intervals were indicated using light emitting diodes. Observers listened monaurally through the left ear, and the increment was assigned to either of the observation intervals with an *a priori* probability of 0.5. Each session consisted of two blocks of 105 trials each. Each block consisted of a single increment duration, and on any one trial the increment could be one of seven amplitudes (15 trials per increment amplitude), with each increment duration associated with a unique set of amplitudes. Eight blocks of trials were collected for each observer at each increment duration.

2.4 Analysis

For each increment duration, an empirical psychometric function was constructed, with percentage correct plotted against increment-to-masker ratio, $\Delta A/A$. Equation (1), with $x=\Delta A/A$ and a scaling constant a, which effectively positions the psychometric function along the x-axis without changing its shape (Irwin, 1989), was fitted to each empirical function. Both k and a were determined by the method of least squares, with goodness-of-fit indexed by the sum of the squared residuals. The best-fitting values of k and a, along with goodness-of-fit statistics, were obtained using model fitting procedures (specifically the Marquardt–Levenberg algorithm) inbuilt into SIGMAPLOT 10.

3. Results

Figure 2 illustrates percentage correct as a function of increment-to-masker ratio, $\Delta A/A$, for five increment durations and demonstrates the dependency of percentage correct upon increment size. Each point on the psychometric function is based on 120 trials. The solid curves are the best-fitting psychometric functions, and were determined by taking the average of the best-fitting parameters a and k for the three observers. Table 1 gives values of k for each observer, along with residual sums of squares, which compare favorably with those reported by Irwin (1989). The scalar a accounts for individual differences in acuity and is not so much of interest as is k, the slope exponent. Reference to Table 1, and examination of Fig. 2, indicates an inverse relationship between increment duration and k. That is, as increment duration increases, k decreases. As a consequence of this relationship, the shape of the psychometric functions become shallower as increment duration increases. This relationship is revealed more clearly in Fig. 3, which illustrates individual values of k as a function of increment duration. As a first approximation, a negative-exponential decline in k occurs as increment duration increases, though individual differences are apparent at 50 and 500 ms durations.

4. Discussion

Figure 3 reveals that the results support the hypothesis that a brief (i.e., 10 ms) increment centered on one of two long-duration maskers constitutes an increment detection task, and not a difference discrimination task. For two observers (HH and DS) the exponent k is close to 1.8 for the 10 ms increment duration, though it is closer to 2.2 for IW. For the shortest two increment

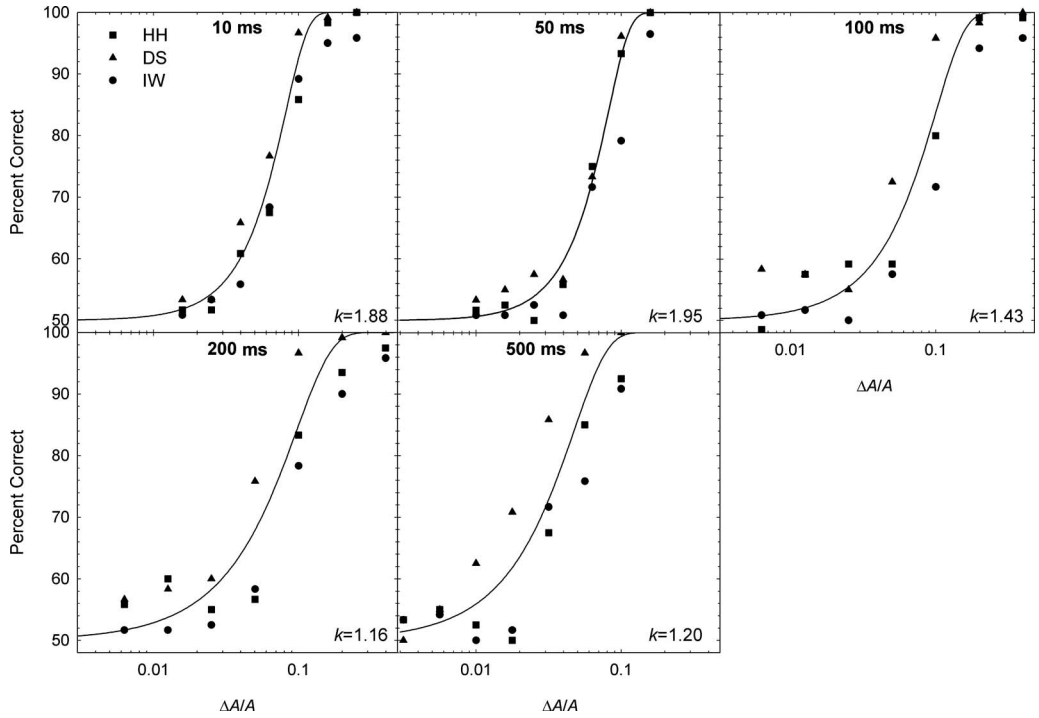


Fig. 2. Data for five increment durations illustrated separately for each observer: HH (■), DS (▲), and IW (●). The solid curves are best-fitting psychometric functions with values of a (not shown) and k (inset) averaged across observers.

durations (10 and 50 ms), a comparison is possible with the data reported by Moore *et al.* (1999), who found a steepening of the psychometric function as increment duration increased from 5 to 50 ms. Figure 3 shows a similar trend for two of the three observers.

At the other end of the increment duration range, 500 ms, support is provided for a difference discrimination task rather than an increment detection task. For two observers (IW and DS), the exponent is close to 1.2, while for HH, the exponent is slightly higher at 1.35. It is worth noting that the latter value results from the worst fit of the predicted psychometric functions to the data (refer Table 1). Notwithstanding this, all three values at the 500 ms increment duration yield 1 when rounded to an integer, and all three values at the 10 ms duration round to 2. This is consistent with the standard predictions made about these two types of task.

Table 1. Best-fitting parameters a and k, and goodness-of-fit indices (SS_{res} and R^2) for three observers and five increment durations.

Increment (ms)	Observer											
	HH				DS				IW			
	a	k	SS_{res}	R^2	a	k	SS_{res}	R^2	a	k	SS_{res}	R^2
10	14.92	1.72	11.51	0.9956	11.16	2.12	43.79	0.9925	9.19	1.79	16.59	0.9786
50	46.53	2.25	29.61	0.9868	28.66	1.55	95.48	0.9779	26.71	2.05	44.71	0.9347
100	20.65	1.31	71.84	0.9110	6.54	1.56	25.65	0.9933	5.52	1.42	86.70	0.9081
200	7.00	1.12	60.94	0.9678	15.81	1.18	58.23	0.9827	5.42	1.18	35.39	0.9687
500	31.01	1.35	114.48	0.9228	55.28	1.12	113.15	0.9595	22.18	1.15	13.38	0.9875

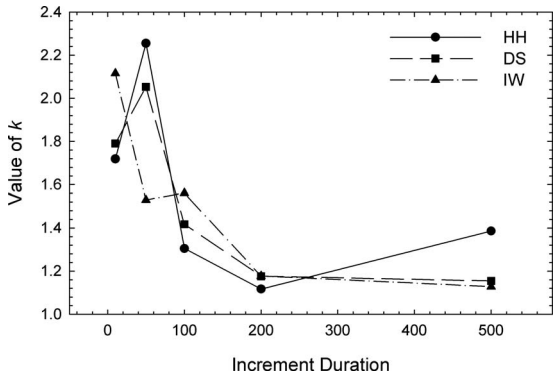


Fig. 3. The slope exponent k as a function of increment duration for three observers.

Of more interest is the relationship between the exponent k and increment duration. With no published data to guide us, we had little basis to predict whether the increment detection task and the difference discrimination task are truly discrete tasks (in which case we might expect a quasi-step function in Fig. 3) or whether the two tasks represent extremes on a continuum with a smooth transition between them for less standard stimulus configurations. Table 1 and Fig. 3 suggest a transition point somewhere between increment durations of 50 and 200 ms. It would be easy to commit the nominal fallacy here. Might the transition between increment detection and difference discrimination be a coalescence: an increment discrimination or difference detection? We are far from convinced that a new category of stimulus configuration with its own unique psychophysical properties would be of much merit, though further investigation is warranted to determine a more exact form of the data presented in Fig. 3. Future studies will need to address the limitations of our study, notably the small number of observers and the small size of the stimulus set.

Results from Moore *et al.* (1999) suggest that increment detection data are more effectively modeled using the Weber fraction expressed in units of intensity (i.e., $\Delta I/I$), while difference discrimination data are better modeled using the level difference, ΔL. Again, with respect to Fig. 3, the question arises as to the nature of the transition between data arising from the difference discrimination task and those arising from the increment detection task. That is, for data of this type, when should $\Delta I/I$ be used as the abscissa and when should this measure be ΔL? Note, however, that such a consideration is less to do with practicalities and more to do with theory. When the increments are as small as those commonly used in level discrimination procedures, it is possible to choose between $\Delta I/I$, $\Delta A/A$, or ΔL with immunity because of the proportional relationship these measures have to one another. The correct measure cannot be determined from data collected using this typical range of levels (Green, 1993; Moore *et al.*, 1999).

Of additional interest is the transition from increment detection to decrement detection. For increment detection, Moore *et al.* (1999) asserted that the correct measure of the stimulus is $\Delta I/I$, whereas for decrement detection the supported measure is ΔL. Consider a task, with two observation intervals, in which a set of increments or decrements are added or subtracted from a continuous pedestal. If the second interval always contains the increment or decrement, and if the task of the observer is to judge if the second interval was more intense than the first, then the task would be the classical method of constant stimuli. On the basis of the conclusions of Moore *et al.* (1999), the x-axis should be divided into two portions at the point of subjective equality (PSE). Stimulus values below the PSE should be measured in ΔL, and those above the PSE in $\Delta I/I$. Considerations such as these emphasize the importance of resolving the measurement problem in level discrimination (Shepherd and Hautus, 2007).

References and links

[1]$\Delta A/A$ is the measure employed here because Laming's (1986) sensory analytical model, in the course of making quantitative predictions of the slope exponent k assumes $\Delta A/A$.

Green, D. M. (**1993**). "Auditory intensity discrimination," in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).
Green, D. M., Nachmias, J., Kearney, J. K., and Jeffress, L. A. (**1979**). "Intensity discrimination with gated and continuous sinusoids," J. Acoust. Soc. Am. **66**, 1051–1056.
Irwin, R. J. (**1989**). "Psychometric functions for the discrimination of differences in intensity of Gaussian noise," Q. J. Exp. Psychol. **41**, 655–674.
Laming, D. (**1986**). *Sensory Analysis* (Academic, London).
McFadden, D. (**1966**). "Masking-level differences with continuous and with burst masking noise," J. Acoust. Soc. Am. **40**, 1414–1419.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1999**). "Effects of frequency and duration on psychometric functions for detection of increments and decrements in sinusoids in noise," J. Acoust. Soc. Am. **106**, 3539–3552.
Shepherd, D., and Hautus, M. J. (**2007**). "The measurement problem in level discrimination," J. Acoust. Soc. Am. **121**, 2158–2167.

Low-frequency attenuation of acoustic waves in sandy/silty marine sediments

Allan D. Pierce and William M. Carey
Department of Mechanical Engineering, Boston University, Boston, Massachusetts 02215
adp@bu.edu

Abstract: Sandy/silty marine sediments are water saturated and consist of diverse tiny rock pebbles. The weight of higher pebbles holds lower pebbles in contact. For low-frequency acoustic disturbances, the no-slip condition and viscosity cause the local water displacement near solid surfaces to be nearly the same as that of the neighboring pebbles. Water farther from surfaces oscillates relative to solid matter because of mass density difference, and viscosity limits the oscillation amplitude. Derived dissipative wave equation predicts attenuation proportional to frequency squared, proportional to the square of the difference of the densities, and inversely proportional to viscosity.

© 2008 Acoustical Society of America
PACS numbers: 43.30.Ma, 43.20.Jr [WS]
Date Received: June 19, 2008 **Date Accepted:** July 12, 2008

1. Introduction

Acoustical models of marine sediments date back to Pekeris,[1] who stated that the "simplest model which exhibits the essential [observed] features…is one in which the bottom is assumed to be a liquid of density…and sound velocity…different from [that in the water column]." In retrospect, this model's success was because the shear wave speed in sediments is much less than the compressional wave (acoustic) speed. It does not in itself offer an explanation for the intrinsic attenuation of waves in the sediment.

A commonly cited continuum-mechanical model, which can account for the low-frequency attenuation in sandy/silty sediments, is the one proposed by Biot,[2] which contains eight "material constants." Later works, such as by Burridge and Keller[3] and Stoll,[4] have confirmed that Biot's model is compatible with fundamental equations governing a medium partly filled with an elastic solid and partly filled with a viscous compressible fluid. Derivable predictions include compressible acoustic waves at low frequencies for which the attenuation α is directly proportional to frequency squared, with a proportionality constant depending on Biot's eight material constants.

The present paper develops a simpler model (and one only slightly more complicated than that of Pekeris) for compressional waves in sandy/silty marine sediments, which is compatible with Biot's model, but which makes use of simplifying aspects of realistic sandy/silty sediments at the outset. The development yields a "physical explanation" and also yields an explicit mathematical prediction of the attenuation in terms of porosity and of the intrinsic properties of water and of the solid particles (here referred to pebbles), which make up the sediment. The development predicts that the resulting model, as well as Biot's original low-frequency model, has applicability for substantially higher frequencies than those suggested by the order-of-magnitude arguments that appear in Biot's paper.

2. Locally averaged stress

The derivation begins with the continuum integral version of Newton's second law, which for a volume V moving such that no there is no local mass transport across the confining surface S is

 © 2008 Acoustical Society of America

$$\frac{d}{dt}\int_V \rho v_i dV = \int_S \sigma_{ij} \boldsymbol{n}_j dS. \tag{1}$$

Here ρ is the local mass density, $\mathbf{v}$ is the local particle velocity, $\boldsymbol{n}$ is the unit vector on the surface pointing out of the volume, and σ_{ij} are the incremental changes in the stress tensor components associated with the acoustic disturbance. (The stress considered here is that which is in excess of the ambient stress associated with gravitationally induced forces and pressures.) The summation convention regarding repeated indices is used.

If the volume V is sufficiently large compared to volumes of size L^3, the integrands on both sides of the above equation may be replaced with local averages, either over small volume elements of order L^3 or over small surface elements of order L^2. Given that a representative pebble "diameter" a is of the order of 10^{-4} m (0.1 mm), a suitable choice for L would be 1 cm. If the sediment has properties that are in an average sense independent of the choice of directions of the coordinate axes, then an average over area should give the same result as an average over volume. Thus the stress components σ_{ij} in the above surface integral are replaced with averages $\langle\sigma_{ij}\rangle$, which are regarded as volume averages. Moreover, if the averaging volume is sufficiently large, the fraction χ_f (porosity) that is fluid should be the same as the fraction for the sediment as a whole, so one writes

$$\sigma_{ij} \rightarrow \sigma_{\mathrm{eff},ij} = \chi_f \langle\sigma_{ij}\rangle_f + \chi_s \langle\sigma_{ij}\rangle_s, \tag{2}$$

where the subscripts, f and s, on the brackets denote the portion, fluid or solid, over which the averaging is to be carried out. The quantity $\chi_s = 1 - \chi_f$ is the fraction of the volume occupied by solid matter. (For sandy/silty marine sediments, the porosity is typically between 0.40 and 0.60.)

For small amplitude disturbances, the time derivative of the volume integral is the integral of the time derivative of the integrand. Thus, with an analogous partitioning of the momentum density average, Eq. (1) becomes

$$\int_V \frac{\partial}{\partial t}[\chi_f \rho_f \langle v_i\rangle_f + \chi_s \rho_s \langle v_i\rangle_s] dV = \int_S [\chi_f \langle\sigma_{ij}\rangle_f + \chi_s \langle\sigma_{ij}\rangle_s] n_j dS. \tag{3}$$

Here the volume V, apart from the restriction that it is large in terms of the averaging length L, is arbitrary, and the integrands are defined so that they are smoothly varying functions of position. Consequently, the chain of reasoning by which Cauchy's equation of motion is customarily derived (via the divergence theorem) yields the partial differential equation

$$\rho_{\mathrm{eff}} \frac{\partial}{\partial t} v_{\mathrm{eff},i} = \frac{\partial}{\partial x_j} \sigma_{\mathrm{eff},ij}, \quad \rho_{\mathrm{eff}} \boldsymbol{v}_{\mathrm{eff}} = \chi_s \rho_s \langle \boldsymbol{v}\rangle_s + \chi_f \rho_f \langle \boldsymbol{v}\rangle_f, \tag{4}$$

where the latter defines the the effective particle velocity. The effective density ρ_{eff} is similarly defined as $\chi_s \rho_s + \chi_f \rho_f$.

3. Mallock–Wood approximation

The type of low-frequency small amplitude disturbances of interest here has wavelengths substantially bigger than L and propagates at a speed roughly comparable to the speed of sound in water (approximately 1500 m/s), with relatively small attenuation. (The highest frequency of interest here is 2000 Hz, which corresponds to a wavelength of the order 0.75 m, and which is still substantially larger than the estimated appropriate choice of 1 cm for L.) For such circumstances, the motion of both the fluid and the solid matter is nearly uniform over distances of order L, and the displacements of the fluid and solid portions are nearly the same, this being ensured by the "no-slip condition" at the interfaces between the pebbles and the water. However, in the limit of zero frequency, the viscosity of the water cannot contribute to the force balance,

so a good first approximation is $\langle \sigma_{ij} \rangle_f \approx -\langle p \rangle_f \delta_{ij}$ where $\langle p \rangle_f$ is the average pressure in the fluid matter. In this low-frequency limit, the local fluid pressure is nearly the same as the average fluid pressure.

Each tiny pebble is almost fully surrounded by water. Contact areas where adjacent pebbles touch should be a small fraction of the total surface area of any given pebble. Also, a valid static solution of the equations for the solid matter, when any portion of it is completely surrounded by water with a constant pressure, is that the stress tensor in the solid is everywhere a constant multiple of the unit tensor and with the diagonal elements equal to the negative of the pressure of the surrounding water.

These two approximations allow one to rewrite Eq. (4) in the simple form

$$\rho_{\text{eff}} \frac{\partial}{\partial t} \boldsymbol{v}_{\text{eff}} = -\boldsymbol{\nabla} \langle p \rangle_f, \tag{5}$$

which is the same as the linearized version of Euler's equation of motion for an inviscid fluid. (Viscosity, however, is not being neglected here—it enters into the formulation at a later stage.)

Increase in pressure acting on any given material volume causes the volume to decrease; the appropriate proportionality factor is the bulk modulus ($B_f = \rho_f c_f^2$ for the fluid and B_s for the solid), so that $\langle p \rangle_f$ is equal to either $-B_f \boldsymbol{\nabla} \cdot \langle \boldsymbol{u} \rangle_f$ or $-B_s \boldsymbol{\nabla} \cdot \langle \boldsymbol{u} \rangle_s$, and so that

$$\langle p \rangle_f = -B_{\text{eff}} [\chi_f \boldsymbol{\nabla} \cdot \langle \boldsymbol{u} \rangle_f + \chi_s \boldsymbol{\nabla} \cdot \langle \boldsymbol{u} \rangle_s], \tag{6}$$

$$B_{\text{eff}} = \left[\frac{\chi_s}{B_s} + \frac{\chi_f}{B_f} \right]^{-1}. \tag{7}$$

The approximations that occur above are, in the essential concepts, due to Mallock,[5] who considered the special case of bubbles in a liquid. The extension to solid particles immersed in a fluid appears in the textbook by Wood.[6] Equations (5) and (6) are incomplete. The Mallock–Wood approximation takes the two displacements to be approximately the same, so that the average pressure satisfies the wave equation

$$\nabla^2 \langle p \rangle_f - \frac{1}{c_{\text{eff}}^2} \frac{\partial^2 \langle p \rangle_f}{\partial t^2} = 0, \quad c_{\text{eff}} = \left(\frac{B_{\text{eff}}}{\rho_{\text{eff}}} \right)^{1/2}, \tag{8}$$

where the latter is the effective sound speed.

4. Imperfect locking and viscosity

A modification to the wave equation that accounts for attenuation uses the definition in Eq. (4) of the effective particle velocity to write

$$\langle \boldsymbol{v} \rangle_f = \boldsymbol{v}_{\text{eff}} + \frac{\rho_s \chi_s}{\rho_{\text{eff}}} \langle \boldsymbol{w} \rangle_f, \quad \langle \boldsymbol{v} \rangle_s = \boldsymbol{v}_{\text{eff}} - \frac{\rho_f \chi_f}{\rho_{\text{eff}}} \langle \boldsymbol{w} \rangle_f. \tag{9}$$

Here $\langle \boldsymbol{w} \rangle_f$ is the difference velocity $\boldsymbol{w} = \boldsymbol{v}_f - \langle \boldsymbol{v} \rangle_s$ averaged over the portion of the volume occupied by the fluid. With substitution of these expressions, Eq. (6) becomes

$$\frac{\partial}{\partial t} \langle p \rangle_f = -B_{\text{eff}} \boldsymbol{\nabla} \cdot \boldsymbol{v}_{\text{eff}} - \frac{B_{\text{eff}} \chi_s \chi_f}{\rho_{\text{eff}}} (\rho_s - \rho_f) \boldsymbol{\nabla} \cdot \langle \boldsymbol{w} \rangle_f. \tag{10}$$

The time derivative of this, followed by a substitution from Eq. (5), yields the corrected wave equation

$$\nabla^2 \langle p \rangle_f - \frac{1}{c_{\text{eff}}^2} \frac{\partial^2}{\partial t^2} \langle p \rangle_f = -\chi_s \chi_f (\rho_s - \rho_f) \frac{\partial}{\partial t} \boldsymbol{\nabla} \cdot \langle \boldsymbol{w} \rangle_f. \tag{11}$$

Derivation of an appropriate approximation for the term on the right of Eq. (11) begins with the equation for fluid motion (velocity $\boldsymbol{v}_f$) in the region occupied by the fluid (η denoting the viscosity),

$$\rho_f \frac{\partial v_{f,i}}{\partial t} = -\frac{\partial}{\partial x_i} p + \eta \frac{\partial}{\partial x_j}\left[\frac{\partial v_{f,i}}{\partial x_j} + \frac{\partial v_{f,j}}{\partial x_i} - \frac{2}{3}(\boldsymbol{\nabla}\cdot\boldsymbol{v}_f)\delta_{ij}\right]. \tag{12}$$

The solid matter velocity field is smooth over distances of the order of L, but the difference in densities plus the no-slip condition imply that there may be a bowing of the fluid velocity profile in the interstices when the overall local region is being accelerated. What counteracts this bowing is the viscosity term in the above.

Isolation of the relevant mechanism is achieved by first approximating the pressure by its spatial average, and also replacing the fluid velocity in the acceleration term with its spatial average. Also, the fluid velocity in the viscous term can be satisfactorily approximated by $\boldsymbol{w} = \boldsymbol{v}_f - \langle \boldsymbol{v}\rangle_s$. The no-slip condition is then approximated by having $\boldsymbol{w}=0$ everywhere on the interfaces. All this leads to

$$\rho_f \frac{\partial}{\partial t}\langle \boldsymbol{v}\rangle_f = -\boldsymbol{\nabla}\langle p\rangle_f + \eta\left[\nabla^2 \boldsymbol{w} + \frac{1}{3}\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{w})\right]. \tag{13}$$

Replacement of the pressure gradient term with the left side of Eq. (5) reduces this to

$$-\chi_s(\rho_s-\rho_f)\frac{\partial \boldsymbol{v}_{\rm eff}}{\partial t} + \frac{\rho_f\rho_s\chi_s}{\rho_{\rm eff}}\frac{\partial}{\partial t}\langle \boldsymbol{w}\rangle_f = \eta\left[\nabla^2\boldsymbol{w} + \frac{1}{3}\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{w})\right]. \tag{14}$$

The formal solution of this for $\boldsymbol{w}$ is written in terms of a dimensionless tensor function W_{ij} as

$$w_i = -W_{ij}(\xi_x,\xi_y,\xi_z)\frac{a^2}{\eta}\left[-\chi_s(\rho_s-\rho_f)\frac{\partial v_{{\rm eff},j}}{\partial t} + \frac{\rho_f\rho_s\chi_s}{\rho_{\rm eff}}\frac{\partial}{\partial t}\langle w_i\rangle_f\right], \tag{15}$$

$$\nabla^2_{\boldsymbol{\xi}} W_{ij} + \frac{1}{3}\frac{\partial^2}{\partial\xi_i\partial\xi_k}W_{kj} = -\delta_{ij}. \tag{16}$$

Here W_{ij} is a function of the scaled coordinates $\xi_x = x/a$, $\xi_y = y/a$, and $\xi_z = z/a$. (The formulation continues to use a for a representative pebble diameter.) The boundary conditions are that each W_{ij} has to be identically zero on all the interfaces, and the domain is the region occupied by the fluid. The scaled coordinates guarantees that each W_{ij} and its derivatives should be dimensionless. Since the sediment is statistically isotropic on the scale of a volume L^3, the averages of the tensor W_{ij} are such that $\langle W_{ij}\rangle_f = \beta\delta_{ij}$ where β is a positive dimensionless number. (As discussed further below, representative experimental data suggest that β is of the order of 0.0013. The small value, relative to unity, is tentatively explained in part as resulting because of the random directions of the unit normal vector $\boldsymbol{n}$ in a representative volume.) Spatial averaging of Eq. (15) over the fluid volume yields

$$\mathcal{L}\langle \boldsymbol{w}\rangle_f = \frac{\beta a^2\chi_s}{\eta}(\rho_s-\rho_f)\frac{\partial \boldsymbol{v}_{\rm eff}}{\partial t} \approx \langle \boldsymbol{w}\rangle_f, \tag{17}$$

where $\mathcal{L}$ is the operator

$$\mathcal{L} = 1 + \frac{\beta a^2\rho_f\rho_s\chi_s}{\rho_{\rm eff}\eta}\frac{\partial}{\partial t} = 1 + \gamma\frac{\partial}{\partial t} \approx 1. \tag{18}$$

The time parameter γ is defined by this equation. Given a value of $\beta = 0.0013$, as is estimated from experimental data further below, the value of γ is estimated to be 1.3×10^{-5} s, which would correspond to a frequency, $f = 1/(2\pi\gamma)$, of 12 kHz. A restriction to frequencies of 2 kHz

or less justifies one's taking the time derivative term in Eq. (18) as negligible, so that the operator $\mathcal{L}$ is set to unity in what follows.

The divergence of Eqs. (17) and (5) yields

$$\nabla \cdot \langle \boldsymbol{w} \rangle_f = \frac{\beta a^2 \chi_s}{\rho_{\text{eff}}} (\rho_s - \rho_f) \nabla^2 \langle p \rangle_f \approx \frac{\beta a^2 \chi_s}{\eta B_{\text{eff}}} (\rho_s - \rho_f) \frac{\partial^2}{\partial t^2} \langle p \rangle_f, \tag{19}$$

where the latter approximation results from the uncorrected wave equation, Eq. (8). The above, when inserted into the corrected wave equation, Eq. (11), yields

$$\nabla^2 \langle p \rangle_f - \frac{1}{c_{\text{eff}}^2} \frac{\partial^2 \langle p \rangle_f}{\partial t^2} = -\frac{\tau}{c_{\text{eff}}^2} \frac{\partial^3 \langle p \rangle_f}{\partial t^3}, \tag{20}$$

where the characteristic time τ is identified as

$$\tau = \frac{\beta a^2 \chi_f \chi_s^2}{\eta \rho_{\text{eff}}} (\rho_s - \rho_f)^2. \tag{21}$$

5. The attenuation coefficient

Insertion of the plane wave function

$$\langle p \rangle_f = A e^{-i\omega t} e^{i[\omega / v_p]x} e^{-\alpha x} \tag{22}$$

into the dissipative wave equation, Eq. (20), gives

$$\frac{\omega}{v_p} + i\alpha = \frac{\omega}{c_{\text{eff}}} [1 + i\omega\tau]^{1/2}, \tag{23}$$

so that in the limit of small angular frequency ω, the attenuation coefficient α (nepers per unit length) is

$$\alpha = \frac{\tau}{2 c_{\text{eff}}} \omega^2 = \frac{\beta a^2 \chi_f \chi_s^2}{2 \eta c_{\text{eff}} \rho_{\text{eff}}} (\rho_s - \rho_f)^2 \omega^2. \tag{24}$$

Explicit calculation of the parameter β from first principles for a realistic sediment is difficult, but order-of-magnitude estimates are roughly consistent with deductions from field experiments. Results reported by Holmes[7] and his colleagues for measurements in Nantucket Sound extrapolate to $\alpha \approx 0.015/8.69$ Np/m at a frequency of 220.5 Hz. (Here 8.69 is the number of decibels in a neper.) For the considered experiments, $\chi_f = 0.49$, $a = 0.115$ mm, $\mu/\rho_f = 10^{-6}$ m^2/s, and $c_{\text{eff}} = 1600$ m/s. With $\rho_s/\rho_f = 2.5$, these numbers lead to $\beta = 0.0013$. This derived value of β and the estimate of the parameter γ given in the preceding section tend to support the claim that the general theory reported here is applicable to marine sediments at frequencies less than 2 kHz.

References and links

[1] C. L. Pekeris, *Theory of Propagation of Explosive Sound in Shallow Water* (Geological Society of America, Boulder, CO, 1948), Memoir 27.
[2] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[3] R. Burridge and J. B. Keller, "Poroelaticity equations derived from microstructure," J. Acoust. Soc. Am. **70**, 1140–1146 (1981).
[4] R. D. Stoll, *Sediment Acoustics* (Springer-Verlag, Berlin, 1989), pp. 5–20.
[5] A. Mallock, "The damping of sound by frothy liquids," Proc. R. Soc. London, Ser. A **84**, 391–395 (1910).
[6] A. B. Wood, *A Textbook of Sound*, 2nd ed. (Macmillan, New York, 1941), pp. 360–362.
[7] J. D. Holmes, "Investigation of Ocean Acoustics using autonomous instrumentation to quantify the water-sediment boundary properties," Ph.D. thesis, College of Engineering, Boston University, Boston, MA, 2007.

Preliminary mapping of void fractions and sound speeds in gassy marine sediments from subbottom profiles

T. G. Leighton
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom
tgl@soton.ac.uk

G. B. N. Robb
National Oceanography Centre, Southampton, European Way, Southampton, SO14 3ZH, United Kingdom
gbor199@noc.soton.ac.uk

Abstract: Bubbles of gas (usually methane) in marine sediments affect the load-bearing properties of the seabed and act as a natural reservoir of "greenhouse" gas. This paper describes a simple method which can be applied to historical and future subbottom profiles to infer bubble void fractions and map the vertical and horizontal distributions of gassy sediments, and the associated sound speed perturbations, even with single-frequency insonification. It operates by identifying horizontal features in the geology and interpreting any perceived change of depth in these as a bubble-mediated change in sound speed.

© 2008 Acoustical Society of America
PACS numbers: 43.30.Ma, 43.20.Hq, 43.30.Pc [AN]
Date Received: June 4, 2008 **Date Accepted:** September 8, 2008

1. Introduction

The bubbles that occur at many seabed locations[1] can impact upon the structural integrity and load-bearing capabilities of the sediment[2,3] and can be indicative of a range of biological, chemical, or geophysical processes[4] (such as global methane budgets). In acoustical terms,[5] such bubbles can degrade the operation of a subbottom profiler (Fig. 1) or be inverted to estimate the bubble population. Bubble radii distributions from 10 μm to 20 mm, and void fractions as great as 9%, have been inferred from the inversion of either compressional wave data,[6–10] acoustic backscatter,[11–13] or two-frequency techniques.[14–16] While such inversions can be based on models of bubbles in water, a model incorporating geotechnical properties of the host sediment surrounding the bubbles (both before and after it is altered by the presence of bubbles) would be preferable.[17]

Table 1 shows that, in addition to the scattering studies,[11–16] there have been many investigations of sound speeds and attenuations of compressional waves in gassy sands and muds, but only one[8] of these has even attempted to estimate void fractions (VFs) remotely (sensitive to VF $>$ 2%). This report outlines a very basic method by which subbottom profiles may be rapidly analyzed to estimate the effect of bubbles on the sound speed in the sediment, and hence map the void fractions, sensitive to VF $=$ 0.002% or better. Furthermore, this method could in principle be applied to single-frequency data. This approach will not replace, but will instead complement, the large-scale field trials which deploy specialist equipment to monitor gas bubbles in sediment. It provides a method either to exploit archived subbottom profiles, or to survey a large area rapidly with commercial equipment from a small vessel.

This paper illustrates the method on a single historical subbottom profile. Supplementary data are not available for this trace, and it is fully recognized that lack of supplementary data on this figure means that the validity of key assumptions (any significant occurrence of severely aspherical bubbles,[18,19] or bubbles whose resonant frequency is not much less than the

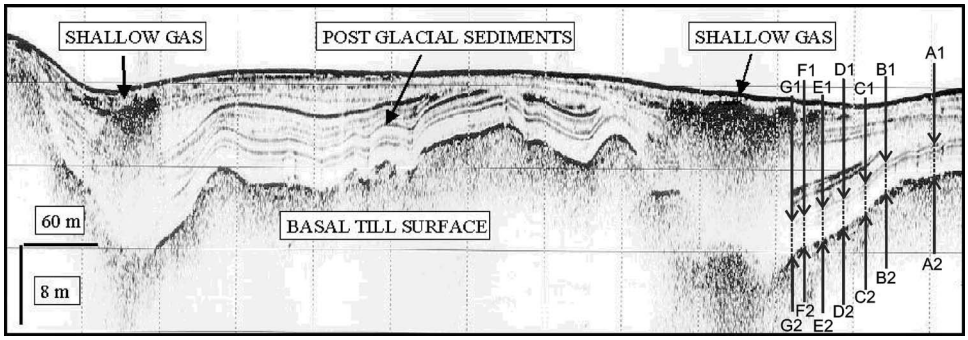


Fig. 1. A chirp subbottom profile from Strangford Lough, Northern Ireland, which displays sediment layers and gas "blanking" (where shallow gas prevents the profiler from obtaining information from beneath the gas). The three interfaces used in this paper to determine sound speed and void fraction are the seabed (upper dark line), the interface denoted by A1 to G1, and the interface denoted by A2 to G2. The chirp signal was a 2–8 kHz linear sweep of 32 ms duration, and the water depth was estimated to be 15.5 m. Reproduced from Ref. 20 and annotated by T. G. Leighton with permission of the National Oceanography Centre, Southampton (J. S. Lenham, J. K. Dix, and J. Bull).

insonifying frequency; the horizontal nature of the sediment layering) cannot be tested without revisiting the site with modern equipment: one purpose of this paper is to show what preliminary estimations can be made from historical data without resorting to such a visit.

2. Method and results

Figure 1 shows a chirp subbottom profile containing gassy sediment.[20] Bubbles cause two "blanking" zones, at which the majority of the sound is scattered (and multiply reflects), and in which no information can be determined on the sediment layering. However, before the geological layering features on either side of the gas pockets become obscured, they appear to dip to greater depths as they approach the "blanking" zones. If a given pair of interfaces in this region was to maintain a constant separation, any apparent changes in separation observed in the subbottom profile can be directly attributed to a change in sound speed in the medium between them.[33,40]

This analysis can be applied to the estimate c_{s-1}, the vertically averaged sound speed between the seabed and interface 1 in Fig. 1 (labeled A1–G1). In a similar way, it can be used to estimate c_{1-2}, the vertically averaged sound speed between interfaces 1 and 2 (labeled A2–G2). The two key assumptions, namely that upper and lower interfaces lie parallel to one another, and that no bubbles exist in the sediment at the location chosen for the calibration (at A), would ideally be justified by the observation of the two layers appearing horizontal over a more extended region than is visible in Fig. 1: no such data were available with this historical record. Recognizing this limitation, the estimated sound speeds perturbations are calculated simply from the ratio of the apparent separation to the assumed separation in Fig. 1 and are plotted in Fig. 2(a).

Retaining the principle that this is an example of using the simplest assumptions to gain a preliminary estimation from this historical data, of the many options[21] for estimating the bubble population from Fig. 2(a) this paper will use Wood's equation[22] for an effective two-fluid medium for illustrative purposes. It is understood that the dynamical response of the bubble population will be influenced by the range of bubble sizes present, and that the bubbles may indeed depart significantly from sphericity. However, assuming that the bubbles are small enough to satisfy the quasistatic conditions inherent in Wood's equation,[21,22] and that the bubbly sediment can be treated as an effective fluid medium, then the effective compressibility is therefore $(1-\beta)(\rho_s c_s^2)^{-1}+\beta(\kappa p_0)^{-1}$, where β is the void fraction of bubbles, ρ_s and c_s are the equilibrium density and sound speed of the bubble-free saturated sediment, p_0 is the static pressure at the position of the bubble, and κ is the polytropic index[17,21] of the gas within the bubbles

Table 1. Compilation of measured compressional wave properties of gassy sediments, including frequency f (kHz), experimental technique, sediment type, compressional wave velocity c_{eff} (m s^{-1}), ratio of velocity in gassy sediment to that in saturated sediment $c_{\mathrm{eff}}/c_{\mathrm{s}}$, attenuation coefficient α (dB m^{-1}), attenuation coefficient of equivalent saturated sediment α_{s} (dB m^{-1}), void fraction, and the source reference. The symbol ··· indicates that no available information is available and the attenuation coefficient of the equivalent sediment was predicted using Hamilton.[31] Note that some simplification and interpretation has been required to condense these data onto a single Table, and that data from sediment containing both free gas and gas hydrate have been omitted from this compilation, owing to the complex four-phase nature of these sediments.[32]

f (kHz)	Site	Sediment type	c_{eff} (m/s)	$c_{\mathrm{eff}}/c_{\mathrm{s}}$	α (dB/m)	α_{s} (dB/m)	VF (%)	Reference no.
···	Remote	Gravel to mud	1210–1650	0.65–0.93	···	···	···	33
···	*In situ*	Mud	110–304	0.06–0.22	···	···	<1	6
0–11	*In situ*	Mud	320–1500	0.21–1.02	<230	<1.3	<6	7
<0.4	Remote	Sand/silt	1250	0.70			<2	8
<1	Remote	Mud	250–1280	0.17–0.86	···	···	···	34
<1	Remote	Mud	550–1100	0.38–0.75	···	···	···	35
<12	Remote	···	800	0.55	···	···	···	36
<20	Remote	···	>1050	0.60			···	37
0.13–1	*In situ*	Clay	800	0.55	1.4/kHz	<0.1/kHz	0.065	9
0.2–3.2	Remote	Mud and sand	45–122	···	···	···	···	38
0.1–2	Lab	Mud/Kaolinite	114–326	0.07–0.21	···	···	0.1–0.4	39
3.5	Remote	Sand to mud	···	0.77			···	40
0.6–1.2	Remote	···	75–170	···	114	···	···	41
1–30	*In situ*	Mud	···	···	2–2433	<4	···	42
3–20	*In situ*	Mud	852–1526	0.56–1.03	···	···	···	43
3–100	Lab	Silty clay	1280	0.88	···	···	<20	44
3.5	Remote	Sand to sandy mud	1610–1660	0.89–0.92	···	···	···	45
5–20	*In situ*	Silty clay	1000–1430	0.70–1.00			0–2	10
7.5	*In situ*	Sand to mud	700–1200	0.47–0.81	···	···	···	46
10–100	Lab	Sand and soil	···	···	2500–7100	<80	···	47
10–1000	Lab	Mud	200–500	0.13–0.34	600–4000	<120	0.4–19.8	48
38	*In situ*	Silty clay	···	···	40–50	<5	0–2	10
40	*In situ*	Mud	···	···	13	<5	···	Cited by 49
40	Lab	Sand	305–1706	0.18–1.00	···	···	···	50
40–80	*In situ*	Mud	···	···	25–90	<10	···	Cited by 49
50	Lab	Glass beads	280–1750	>0.16	···	···	0.1–100	51
50	Lab	Sand	250–1700	>0.15	···	···	0.1–100	51
110	*In situ*	Soil	1220–1270	0.84–0.87	105–470	···	···	52
200	Lab	Sand	1218–2090	0.58–1.00	···	···	0–100 controlled	53 and 54
300–700	Lab	Mud	1400–1700	0.95–1.15	<700	<84	<6	7
400	Lab	Mud	1522–1523	1.00	···	···	0.1–0.4	39
400	Lab	Silty clay	1430	1.00	>500	<48	0–2	10
400	Lab	···	736–1372	0.50–0.94	···	···	···	55
400	Lab	Mud	1200–1300	0.80–0.87	···	···	···	56
500	Lab	Sand to mud	1900–1460	0.62–1.00	···	···	···	57
500–1000	Lab	Sand and silt	<1000	<0.56	···	···	···	58
700	Lab	Sand	1264–2515	···	···	···	0–100 controlled	59
1000	Lab	Sand to mud	1300	0.88	···	···	···	46

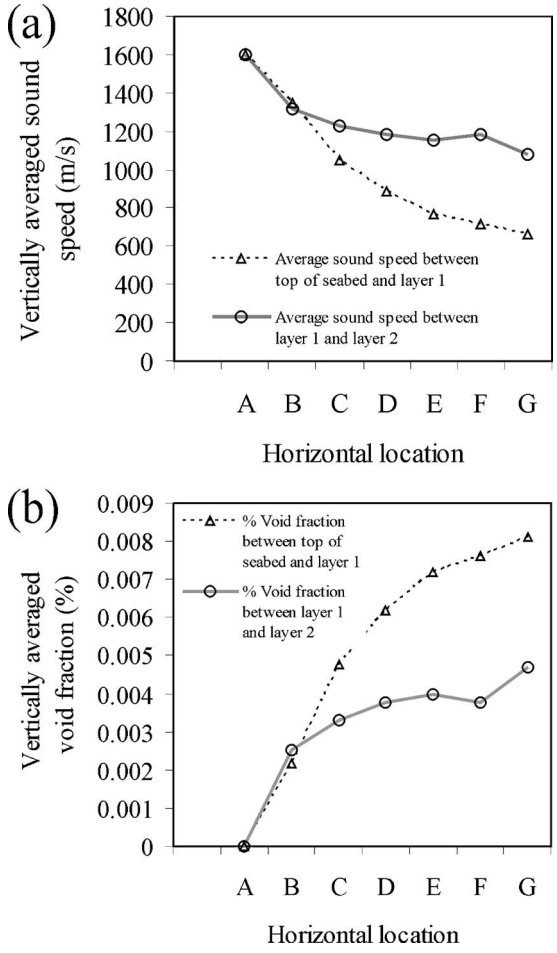


Fig. 2. (a) The vertically averaged sound speeds between the seabed and interface 1, and between interface 1 and 2; (b) the vertically averaged void fractions between the seabed and interface 1, and between interface 1 and 2.

(equal to the ratio of the specific heats of the gas, divided by the polytropic coefficient[7,10,12]), which is assumed to be 1.3 for nearly adiabatic pulsations of biogenic sedimentary gas.[9] The void fraction being small, $\beta \ll 1$, means that the bubbly sediment can be treated as an effective fluid medium of density ρ_{eff}, where $\rho_{\text{eff}}^{-1} \approx \rho_{\text{s}}^{-1}$, since if $\rho_{\text{eff}} = \rho_{\text{s}}(1+\Delta\rho/\rho_{\text{s}})$, then $\rho_{\text{eff}}^{-1} \approx \rho_{\text{s}}^{-1}(1-\Delta\rho/\rho_{\text{s}}+\cdots) \Rightarrow \Delta\rho/\rho_{\text{eff}} \approx \Delta\rho/\rho_{\text{s}} - (\Delta\rho/\rho_{\text{s}})^2 + \cdots$ if $\Delta\rho/\rho_{\text{s}} \ll 1$. Equating the above effective compressibility with $(\rho_{\text{eff}} c_{\text{eff}}^2)^{-1} \approx (\rho_{\text{s}} c_{\text{eff}}^2)^{-1}$ (where c_{eff} is the effective sound speed of the gassy sediment, noting that dissipation is neglected in this definition[21]), implies $\beta = (c_{\text{s}}^2/c_{\text{eff}}^2 - 1)[\rho_{\text{s}} c_{\text{s}}^2/(\kappa p_0) - 1]^{-1}$. Assuming $\kappa p_0 \ll \rho_{\text{s}} c_{\text{s}}^2$ and $|c_{\text{s}}^2/c_{\text{eff}}^2 - 1| \ll 1$, this reduces to

$$\beta \approx \frac{2\kappa p_0}{c_{\text{s}}^2 \rho_{\text{s}}} \left(1 - \frac{c_{\text{eff}}}{c_{\text{s}}}\right). \tag{1}$$

A detailed form of this basic derivation may be found in Ref. 23, and more sophisticated versions may be used if the above assumptions are violated.[17,21] Application of Eq. (1) to the data in Fig. 2(a) allows estimation of the vertically averaged void fraction between the top of the seabed and layer 1 ($\beta_{\text{s}-1}$), and the vertically averaged void fraction between layer 1 and layer 2 (β_{1-2}). At the time the measurements of Fig. 1 were taken, techniques for measuring the density and

sound speed in the sediment were not as advanced as they are today. These values are therefore predicted for the silts and clays typical of this section of Strangford Lough through the use of empirical regressions,[24] resulting in a bubble-free sound speed of 1600 m s^{-1} and a bubble-free density of 2300 kg m^{-3}. Nowadays it is commonplace to examine cores from regions of saturated sediment to determine density from gamma ray attenuation and compressional wave velocity from ultrasonic propagation at 500 kHz (which for these saturated sediments is assumed to provide a sound speed relevant to the frequencies used in this paper).[25]

The final parameter required for the calculation is the pressure p_0. This was computed for the central portion of each layer using $p_0 = p_A + \rho_w g h_w + \rho_s g h/2$, where p_A is the atmospheric pressure on the survey date (103 kPa), g is the acceleration due to gravity, ρ_w is the density of water (1000 kg m^{-3}), h_w is the depth of the water (15.5 m), and h is the depth of the sediment layer below the seabed (5.1 m for layer 1, and 7.4 m for layer 2). This resulted in an average hydrostatic pressure of 312.5 kPa for layer 1 and 338.5 kPa for layer 2.

Figure 2 summarizes the sound speeds and void fractions calculated for each of the coordinates A to G. The assumptions in the sediment parameters introduce an unknown uncertainty.

3. Discussion and conclusion

This report outlines a simple scheme for assessing the sound speed perturbation induced by bubbles in the seabed, and for estimating the void fraction and extent of the bubble layers (including changes both in the horizontal and vertical). Although in the present work this technique has only been applied to two sediment layers (three interfaces), the technique can be extended to include more interfaces and even three-dimensional profiles.[26] Contrast enhancement[27] of the geological layers compared to the bubble scatterers will extend the technique to higher void fractions, closer to the center of the gas pocket.

The approach provides a quick first-order technique, but the simplicity of its use is offset by limitations. The conditions may violate key assumptions (quasi-static dynamics of noninteracting spherical bubbles), and inhomogeneities may contradict the assumed averaging. Nevertheless, the ease with which first-order environmental data can be gained at little extra effort using existing technology, and through examination of historical records of subbottom profiles, offers the possibility of making rapid progress. This is significant given

(i) quantitative sound speeds and void fractions can be estimated from subbottom profiles which were previously used qualitatively to assess the location of gassy sediments;
(ii) the method shows that the extent of the gas, as indicated by the void fractions shown in Fig. 2(b), is much greater than the extent of the shadows in Fig. 1; which one might otherwise consider to represent the location and extent of the gas pocket;
(iii) the method can be implemented without the complex equipment often required to assess gassy sediments (e.g., difference frequency sonars, CT scanners, etc.);[10,28–30]; and
(iv) the method is remote and highly sensitive (compare with Table 1).

This method is not presented as an alternative to the more sophisticated methods under development, or as competition for the innovative and large-scale field trials specifically designed to measure *in situ* bubble populations in sediments. Instead it is envisaged to be a complementary technique which would allow the rapid low-cost assessment of a gassy area using current commercial apparatus, and the new analysis of historical data. Geological expertise will be required in each case to assess the likelihood that the interfaces selected are parallel, and that the perceived dipping is solely due to sound speed perturbations[33,37,40] (though, because gassy sediments are dispersive, this can be tested remotely by taking additional profiles of the same location at higher frequencies and seeing whether the apparent separations remain constant). In the model used here, the material parameters of the sediment enter only through the sound speed and density of saturated sediment, and there is therefore no reflection of the complexity of propagation that can occur in such materials. While this simplification could be overcome through the substitution of improved models for sound speed into this scheme[17] (and

while the assumption of quasi-static dynamics can be replaced using a more sophisticated inversion routine), the importance of this report lies in expressing such a simple scheme for obtaining the void fraction and extent (in the vertical and horizontal) of bubble populations in marine sediment. It can also be applied to tissue, and to domestic products and pharmaceuticals into which deliberate target layers could be placed.

Acknowledgments

This work is funded by the Engineering and Physical Sciences Research Council, Grant No. EP/D000580/1 (Principal Investigator: T.G. Leighton). The authors are grateful to Justin Dix for providing the background data to Fig. 1 (water depth, and the density and sound speed of the bubble-free sediment) and to Peter Birkin for helpful comment.

References and links

[1]P. Fleischer, T. H. Orsi, M. D. Richardson, and A. L. Anderson, "Distribution of free gas in marine sediments: A global overview," Geo-Mar. Lett. **21**, 103–122 (2001).
[2]S. J. Wheeler and T. N. Gardiner, "Elastic moduli of soils containing large gas bubbles," Geotechnique **39**, 333–342 (1989).
[3]G. C. Sills, S. J. Wheeler, S. D. Thomas, and T. N. Gardiner, "Behaviour of offshore soils containing gas bubbles," Geotechnique **41**, 227–241 (1991).
[4]A. G. Judd, "The global importance and context of methane escape from the seabed," Geo-Mar. Lett. **23**, 147–154 (2003).
[5]A. G. Judd and M. Hovland, "The evidence of shallow gas in marine sediments," Cont. Shelf Res. **12**, 1081–1095 (1992).
[6]A. R. Tinkle, K. R. Wener, and C. A. Meeder, "Seismic no-data zone, offshore 1988, Mississippi delta: Part 1-Acoustic characterisation," Proceedings to the 20th Offshore Technology Conference (Houston) (1988) pp. 65–74.
[7]A. I. Best, M. D. J. Tuffin, J. K. Dix, and J. M. Bull, "Tidal height and frequency dependence of acoustic velocity and attenuation in shallow gassy marine sediments," J. Geophys. Res. **109**, B08101 (2004).
[8]K. Andreassen, P. E. Hart, and M. MacKay, "Amplitude versus offset modelling of the bottom stimulated reflection associated with submarine gas hydrates," Mar. Geol. **137**, 25–40 (1997).
[9]T. S. Edrington and T. M. Calloway, "Sound speed and attenuation measurements in gassy sediments in the Gulf of Mexico," Geophysics **49**, 297–299 (1984).
[10]R. H. Wilkens and M. D. Richardson, "The influence of gas bubbles on sediment acoustic properties: In situ, laboratory, and theoretical results from Eckernförde Bay, Baltic Sea," Cont. Shelf Res. **18**, 1859–1892 (1998).
[11]F. A. Boyle and N. P. Chotiros, "A model for high-frequency acoustic backscatter from gas bubbles in sandy sediments at shallow grazing angles," J. Acoust. Soc. Am. **98**, 531–541 (1995).
[12]L. Fonseca, L. Mayer, D. Orange, and N. Driscoll, "The high-frequency backscattering angular response of gassy sediments: Model/data comparison from the Eel River margin, California," J. Acoust. Soc. Am. **111**, 2621–2631 (2002).
[13]A. P. Lyons, M. E. Duncan, A. L. Anderson, and J. A. Hawkins, "Predictions of the acoustic scattering response of free-methane bubbles in muddy sediments," J. Acoust. Soc. Am. **99**, 163–172 (1996).
[14]Z. Klusek, A. Sutin, A. Matveev, and A. Potapov, "Observation of nonlinear scattering of acoustical waves at sea sediments," Acoust. Lett. **18**, 198–203 (1995).
[15]S. V. Karpov, Z. Klusek, A. L. Matveev, A. I. Potapov, and A. M. Sutin, "Nonlinear interaction of acoustic wave in gas-saturated marine sediments," Acoust. Phys. **42**, 464–470 (1996).
[16]J. Tegowski, Z. Klusek, and J. Jaromir, "Nonlinear acoustical methods in the detection of gassy sediments," in *Acoustic Sensing Techniques for the Shallow Water Environment*, edited by A. Caiti, N. R. Chapman, J.-P. Herman, and S. M. Jesus (Springer, Berlin, 2006), pp. 125–136.
[17]T. G. Leighton, "Theory for acoustic propagation in marine sediment containing gas bubbles which may pulsate in a non-stationary nonlinear manner," Geophys. Res. Lett. **34**, L17607 (2007).
[18]K. M. Dorgan, S. R. Arwade, and P. A. Jumars, "Burrowing in marine muds by crack propagation: Kinematics and forces," J. Exp. Biol. **210**, 4198–4212 (2007).
[19]A. H. Reed, B. P. Boudreau, C. Algar, and Y. Furukawa, "Morphology of gas bubbles in mud: A microcomputed tomographic evaluation," in *Proc. 1st Int. Conf. Underwater Acoustic Measurements: Technologies and Results (Heraklion)* (2005), pp. 103–110.
[20]J. M. Lenham, J. M. Bull, and J. K. Dix, "A marine geophysical survey of Strangford Lough," Archaeol. Irel. **11**, 18–20 (1998).
[21]T. G. Leighton, S. D. Meers, and P. R. White, "Propagation through nonlinear time-dependent bubble clouds and the estimation of bubble populations from measured acoustic characteristics," Proc. R. Soc. London, Ser. A **460**(2049), 2521–2550 (2004).
[22]A. Mallock, "The damping of sound by frothy liquids," Proc. R. Soc. London, Ser. A **84**(572), 391–395 (1910).
[23]T. G. Leighton, "A method for estimating sound speed and the void fraction of bubbles from sub-bottom

sonar images of gassy seabeds," ISVR Technical Report No. 320 (2007).
[24]M. D. Richardson and K. B. Briggs, "On the use of acoustic impedance values to determine sediment properties," Proc. Institute of Acoustics **15**, 15–23 (1993).
[25]G. B. N. Robb, A. I. Best, J. K. Dix, J. Bull, T. G. Leighton, and P. R. White, "The frequency-dependence of compressional wave velocity and attenuation coefficient in marine sediment," J. Acoust. Soc. Am. **120**(5), 2526–2537 (2006).
[26]M. Gutowski, J. M. Bull, J. K. Dix, T. J. Henstock, P. Hogarth, T. Hiller, T. G. Leighton, and P. R. White, "Three-dimensional high-resolution acoustic imaging of the sub-seabed," Appl. Acoust. **69**(5), 412–421 (2008).
[27]T. G. Leighton, D. C. Finfer, and P. R. White, "Experimental evidence for enhanced target detection by sonar in bubbly water," Hydroacoustics **11**, 181–202 (2008).
[28]L. A. Ostrovsky, S. N. Gurbatov, and J. N. Didenkulov, "Nonlinear acoustics in Nizhni Novgorod (A review)," Acoust. Phys. **51**, 114–127 (2005).
[29]A. L. Anderson, F. Abegg, J. A. Hawkins, M. E. Duncan, and A. P. Lyons, "Bubble populations and acoustic interaction with the gassy floor of Eckernforde Bay," Cont. Shelf Res. **18**, 1807–1838 (1998).
[30]B. P. Boudreau, C. Algar, B. D. Johnson, I. Croudace, A. Reed, Y. Furukawa, K. M. Dorgan, P. A. Jumars, A. S. Grader, and B. S. Gardiner, "Bubble growth and rise in soft sediments," Geology **33**, 517–520 (2005).
[31]E. L. Hamilton, "Compressional-wave attenuation in marine sediments," Geophysics **37**(4), 620–646 (1972).
[32]S. Chand, T. A. Minshull, J. A. Priest, A. I. Best, C. R. I. Clayton, and W. F. Waite, "An effective medium inversion algorithm for gas hydrate quantification and its application to laboratory and borehole measurements of gas hydrate-bearing sediments," Geophys. J. Int. **166**, 543–552 (2006).
[33]M. L. Holmes and D. L. Thor, "Distribution of gas-charged sediment in Norton Sound and Chirkov Basin, Northeastern Bering Sea," Geol. Mijnbouw **61**, 79–89 (1982).
[34]F. K. Levin, "The seismic properties of Lake Maracaibo," Geophysics **27**, 35–47 (1962).
[35]M. P. Hochstein, "Seismic measurements in Suva Harbour (FIJI)," N.Z. J. Geol. Geophys. **13**, 269–281 (1970).
[36]G. H. Lee, D. C. Kim, H. J. Kim, H. T. Jou, Y. J. Lee, and S. C. Park, "Shallow gas off the southeastern coast of Korea," Proc. VIII Conf. Gas in Mar. Sed. (Vigo), 2005, pp. 87–89.
[37]P. Hempel, V. Spiess, and R. Schreiber, "Expulsion of shallow gas in the Skagerrak-Evidence from subbottom profiling, seismic, hydroacoustic and geochemical data," Estuarine Coastal Shelf Sci. **38**, 583–601 (1994).
[38]J. L. Jones, C. B. Leslie, and L. E. Barton, "Acoustic characteristics of underwater bottoms," J. Acoust. Soc. Am. **36**, 154–157 (1964).
[39]P. S.Wilson, A. H. Reed, W. T. Wood, and R. A. Roy, "Low frequency sound speed measurements paired with computed X-Ray tomography imaging in gas bearing reconstituted natural sediments," Proc. 2nd Int. Conf. Und. Ac. Meas. (Heraklion), 2007, pp. 21–29.
[40]B. S. Hart and T. S. Hamilton, "High-resolution acoustic mapping of shallow gas in unconsolidated sediments beneath the strait of Georgia, British Columbia," Geo-Mar. Lett. **13**, 49–55 (1993).
[41]J. L. Jones, C. B. Leslie, and L. E. Barton, "Acoustic characteristic of a lake bottom," J. Acoust. Soc. Am. **30**, 142–145 (1958).
[42] A. B.Wood and D. E. Weston, "The propagation of sound in mud," Acustica **14**, 156–162 (1964).
[43]S. S. Fu, R. H. Wilkens, and L. N. Frazer, "In situ velocity profiles in gassy sediments: Kiel Bay," Geo-Mar. Lett. **16**, 249–253 (1996).
[44]P. E. Kepkay and R. C. Cooke, "Velocity of sound as a function of bubble distribution in gas-bearing sediments," Geophys. Res. Lett. **5**, 1071–1073 (1978).
[45]S. K. Addy and J. L. Worzel, "Gas seeps and subsurface structure off Panama City, Florida," AAPG Bull. **63**, 668–675 (1979).
[46]T. J. Gorgas, G. Y. Kim, S. C. Park, R. H. Wilkens, D. C. Kim, G. H. Lee, and Y. K. Seo, "Evidence for gassy sediments on the inner shelf of SE Korea from geoacoustic properties," Cont. Shelf Res. **23**, 812–834 (2003).
[47]W. L. Nyborg, I. Rudnick, and H. K. Schilling, "Experiments on acoustic absorption in sand and soil," J. Acoust. Soc. Am. **22**, 422–425 (1950).
[48]T. N. Gardiner, "An acoustic study of soils that model seabed sediments containing gas bubbles," J. Acoust. Soc. Am. **107**, 163–176 (2000).
[49]A. L. Anderson and L. D. Hampton, "Acoustics of gas bearing sediments I, Background," J. Acoust. Soc. Am. **67**, 1865–1889 (1980).
[50]H. Brandt, "Factors affecting compressional wave velocities in unconsolidated marine sand sediments," J. Acoust. Soc. Am. **32**, 171–179 (1960).
[51]M. Kimura, "Acoustic characteristics of marine sediment model with partial gas saturation," Jpn. J. Appl. Phys., Part 1 **31**, 111–114 (1991).
[52]L. D. Hampton and A. L. Anderson, "Acoustics and gas in sediments: Applied Research Laboratory (ARL) experience," in *Natural Gas in Marine Sediments*, edited by I. R. Kaplan (Plenum Press, New York, 1974), pp. 249–273.
[53]S. N. Domenico, "Effect of brine-gas mixture on velocity in an unconsolidated sand reservoir," Geophysics **4**, 882–894 (1976).
[54]S. N. Domenico, "Elastic properties of unconsolidated porous sand reservoirs," Geophysics **42**, 1339–1368 (1977).
[55]D. C. Kim, J. Y. Yeo, G. H. Lee, and S. C. Park, "Seismic characteristics and physical properties of gas-bearing sediment in the Jinhae Bay, the South Sea of Korea," Proc. VIII Conf. Gas in Mar. Sed. (Vigo), 2005, pp.

185–190.
[56]F. Yuan, J. D. Bennell, and A. M. Davis, "Acoustic and physical characteristics of gassy sediments in the western Irish Sea," Cont. Shelf Res. **12**, 1121–1134 (1992).
[57]N. C. Slowey, W. R. Bryant, and D. N. Lambert, "Comparison of high-resolution seismic profiles and the geoacoustic properties of Eckernforde Bay sediments," Geo-Mar. Lett. **16**, 240–248 (1996).
[58]W. J. Winters, W. F. Waite, D. H. Mason, L. Y. Gilbert, and I. A. Pecher, "Methane gas hydrate effect on sediment acoustic and strength properties," J. Pet. Sci. Eng. **56**, 127–135 (2007).
[59]S. E. Elliot and B. F. Wiley, "Compressional velocities of partially saturated unconsolidated sands," Geophysics **40**, 949–954 (1975).

The effect of signal-temporal uncertainty on detection in bursts of noise or a random-frequency complex

Angela Yarnell Bonino and Lori J. Leibold
Department of Allied Health Sciences, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599
abonino@med.unc.edu, leibold@med.unc.edu

Abstract: This study examined the effect of signal-temporal uncertainty on detection of a 120-ms, 1-kHz tone in the presence of a continuous sequence of 120-ms bursts of either a broadband noise or a random-frequency, two-tone complex. Using the method of constant stimuli, signal-temporal uncertainty was defined as the difference in threshold across temporally uncertain and temporally defined listening conditions. Results indicted an average effect of signal-temporal uncertainty of 2 dB for the noise masker compared to 9 dB for the random-frequency, two-tone masker. These results suggest that signal-temporal uncertainty may be more detrimental for conditions in which informational masking dominates performance.

© 2008 Acoustical Society of America
PACS numbers: 43.66.Dc, 43.66.Lj [QJF]
Date Received: July 16, 2008 **Date Accepted:** September 4, 2008

1. Introduction

Multiple sources of acoustic uncertainty are encountered in natural environments. Relevant signals can occur at unpredictable times, and listeners do not always know what spectral and level information will be contained in the signals when they occur. In the laboratory, stimulus uncertainty is typically introduced by varying dimensions of the target signal, the background masker, or both. Large detrimental effects of stimulus uncertainty have been observed, particularly for masker-spectral uncertainty. In this context, masker-spectral uncertainty is produced by randomly varying the frequency components of a multi-tonal masker on each presentation. Compared to performance in quiet, masker-spectral uncertainty can elevate detection thresholds for a fixed-frequency tone by as much as 50 dB (e.g., Neff and Green, 1987). Moreover, substantial masking can occur in the absence of masker energy falling within a presumed auditory filter around the signal. The term "informational" masking is often used to distinguish these effects from peripherally based or "energetic" masking, although factors other than stimulus uncertainty can also produce substantial informational masking (reviewed by Durlach, 2006).

Several investigations have examined the masking produced by randomly varying the spectral content of a multi-tonal masker (e.g., Neff and Green, 1987; Kidd *et al.*, 1994), but effects of other types of uncertainty have received less investigation. Researchers have noted, however, that stimulus uncertainty in dimensions other than frequency and/or the combination of uncertainty across multiple dimensions can contribute additional detriments to subjects' performance (e.g., Durlach *et al.*, 2003). The purpose of the current study was to examine the effect of one potential contributor to informational masking, signal-temporal uncertainty, in the context of a multi-tonal informational masking paradigm.

Previous studies of pure-tone detection in quiet or in noise have found that subjects' performance improves when signals occur at expected versus unexpected times (e.g., Egan *et al.*, 1961; Chang and Viemeister, 1991). Furthermore, subjects are better able to focus attention on auditory events when they are predictable in time (reviewed by Jones and Yee, 1993). Nonetheless, the effect of signal-temporal uncertainty appears to be small for conditions in

which energy-based masking dominates performance (e.g., Egan *et al.*, 1961; Watson and Nichols, 1976; Green and Weber, 1980). For example, Green and Weber (1980) varied the temporal position of a 1-kHz sinusoid in a sequence of ten, 50-ms bursts of broadband-noise. The effect of signal-temporal uncertainty was less than 2 dB.

In contrast to studies of tone-in-noise detection, the effect of signal-temporal uncertainty may be larger for conditions that produce substantial informational masking. Leibold and Werner (2006) examined infants' detection of a 1-kHz tone in the presence of repeating presentations of a random two-tone masker. Thresholds for adult control subjects were higher than previously reported for stimuli with similar spectral and level properties (e.g., Neff and Dethlefs, 1995). One explanation for the elevated thresholds reported by Leibold and Werner is that procedural modifications designed to facilitate infant testing increased signal-temporal uncertainty. In contrast to studies that have examined informational masking in adults using 2IFC or yes/no tasks in which the listening interval was defined, subjects did not have *a priori* knowledge regarding the onset of signals. Thus, subjects were listening under conditions of both masker-spectral and signal-temporal uncertainty.

Recent studies are consistent with the idea that the effect of signal-temporal uncertainty depends on the particular stimuli and procedures employed. Best *et al.* (2007) found that subjects could more accurately identify birdsong when provided with a visual cue defining the time interval that contained the target. However, a smaller improvement was observed for recalling digits presented in time-reversed speech. Best *et al.* suggested that the benefit of knowing when to listen depends on the salience of the stimuli and the listener's experience with them. In work related to their infant study, Leibold and Werner (2003) also suggest that the benefit of knowing when to listen is larger for conditions in which informational masking dominates performance. Detection thresholds for a 1-kHz tone were estimated in the presence of a random two-tone masker. The average threshold for five trained adults was 16 dB lower when the listening interval was defined by a visual cue compared to when the listening interval was uncertain. However, response bias was not accounted for with the single-interval adaptive procedure. Thus, it is difficult to determine whether this threshold difference reflects a difference in sensitivity or a difference in detection strategy.

Based on Leibold and Werner (2003), the current study examined whether the effect of signal-temporal uncertainty is larger for conditions that generate considerable informational masking. Using the method of constant stimuli, thresholds for a 120-ms, 1-kHz tone were computed from an unbiased estimate of sensitivity (d') in the presence of 120-ms bursts of either a broadband noise or a random-frequency, two-tone complex. For each masker, the effect of signal-temporal uncertainty was defined as the threshold difference across temporally uncertain and temporally defined conditions. The hypothesis was that the effect of signal-temporal uncertainty would be larger for the random two-tone masker, expected to produce substantial informational masking, than for the noise masker.

2. Method

2.1 Subjects

The subjects were four adults (21 to 27 years) with thresholds in quiet of $\leqslant$20 dB HL for octave frequencies from 0.25 to 8 kHz, bilaterally. Subjects were tested individually in a double-walled, sound-treated room for 2-h sessions, including regular breaks. All subjects had experience with both the task and stimuli and were selected because they exhibited substantial masking in the presence of random two-tone maskers. An additional subject was tested but excluded from analysis. Despite multiple test blocks, this subject's data remained variable and resulted in a poorly fitted psychometric function for one condition ($r^2=0.4$).

2.2 Stimuli and equipment

In all conditions, the signal was a 1-kHz pure tone of 120-ms total duration, including 5-ms, $\cos^2$ onset/offset ramps. There were two types of maskers: broadband noise (0.3–3 kHz) and

random-frequency, two-tone complexes. Maskers were presented at an overall level of 50 dB SPL (47 dB/component) for a total duration of 120-ms, including 5-ms, $\cos^2$ onset/offset ramps. The 120-ms masker bursts repeated continuously throughout the testing session with no temporal overlap between successive masker bursts. For the random two-tone masker, frequency components for each 120-ms burst were independently selected from two different uniform distributions. One component was drawn from a range of 0.3–0.92 kHz and the second from a range of 1.08–3 kHz. The 1-kHz signal, when present, was gated simultaneously with one of the repeating 120-ms masker bursts.

Stimuli were generated digitally at a 25-kHz sampling rate using Tucker-Davis Technologies (TDT) system III programmable hardware (RP2) and then low pass filtered at 4 kHz. Signal and masker were attenuated separately (TDT PA5), mixed, fed to a headphone buffer (TDT HB7), and to the left ear of the subject through an insert phone (ER1, Etymotic Research). The presentation of stimuli was controlled by a computer using custom software.

2.3 Procedure

Across four conditions, two experimental variables were manipulated: (1) masker type (noise versus random two-tone) and (2) signal-temporal uncertainty (uncertain versus defined listening interval). Trials were conducted in runs of 20 trials using a single-interval, yes/no procedure. For temporally uncertain conditions, an observer outside the booth initiated each trial. Trials were either "signals", in which the 1-kHz tone was presented simultaneously with one presentation of the repeating masker, or "nonsignals" in which only the masker was presented. When present, the signal coincided with the second masker burst following the observer's initiation of the trial. Subjects were not provided information regarding when an observation interval occurred and were instructed to press ⟨+⟩ on a keyboard whenever they heard the 1-kHz tone. The listening interval ended when the subject indicated a response or after 4 s (maximum of 33 masker bursts). A failure to press ⟨+⟩ during the 4-s listening interval indicated a "no" response. Trial-by-trial feedback was not provided, but subjects were informed of the overall hit and false alarm rate following each run of 20 trials.

An identical procedure was used for temporally defined conditions, except that the subject (rather than the observer outside of the booth) initiated each trial by pressing ⟨enter⟩ on a keyboard. In contrast to Best *et al.* (2007) and Leibold and Werner (2003), visual cues were not used to indicate the listening interval. Instead, the listening interval was predictable and coincided with the second masker burst following the subject's initiation of each trial.

Subjects had at least 2 h of training with the random two-tone masker for both temporal conditions prior to testing. After completing training, data were collected in runs of 20 trials, with signal level fixed within a run. Five fixed signal levels were selected for each subject and condition. The mid-point of the five levels was estimated using a one-up, one-down, single-interval adaptive procedure that tracked the 50% point on the psychometric function (Levitt, 1971). The remaining four signal levels corresponded to −6, −3, +3, and +6 dB around the mid-point. This approach ensured an individualized range of signal levels spanning the dynamic range of the psychometric function for each subject.

For testing, a constant-stimuli procedure with a 0.5 *a priori* probability of signals was used. Subjects were informed of the signal-nonsignal probability. Masker type and temporal condition were fixed within a given block of runs. Five runs were conducted for each block (100 trials total), varying signal level across runs. Testing order for fixed-level runs was counterbalanced across subjects, with subjects completing all conditions for a given masker before changing masker type. For each condition, testing was conducted in a descending order of signal level. The entire sequence was then repeated in an ascending order of signal level, resulting in 40 trials at each of the five signal levels for each condition.

For each subject and condition, values of d' were computed for the five signal levels and used to construct a psychometric function following Dai (1995). This procedure estimates a

Table 1. Individual and mean estimates of threshold (dB SPL), slope (log d'/dB) and criterion from the individual fitted psychometric functions. Values of r^2 show the proportion of variance in the data accounted for by the estimates.

	Noise masker							
	Uncertain				Defined			
Subject	Threshold	Slope	Criterion	r^2	Threshold	Slope	Criterion	r^2
S1	33.3	0.7	1.0	0.88	34.6	1.3	0.6	0.90
S2	35.5	0.8	0.6	0.92	32.4	0.8	0.7	0.95
S3	34.9	1.1	1.4	0.94	33.8	1.0	1.1	0.90
S4	34.6	0.9	1.5	0.83	30.0	0.5	1.0	0.98
Mean	34.6	0.9	1.1	0.89	32.7	0.9	0.9	0.93
	Random two-tone masker							
S1	47.5	0.4	0.1	0.67	32.2	0.3	0.3	0.57
S2	17.2	0.2	0.6	0.51	5.2	0.2	0.8	0.86
S3	18.3	0.3	0.8	0.77	13.3	0.4	0.6	0.54
S4	30.3	0.2	0.5	0.58	25.3	0.6	0.2	0.76
Mean	28.3	0.3	0.5	0.63	19.0	0.4	0.5	0.68

threshold corresponding to $d'=1$. Psychometric functions were fitted based on 200 trials, with 40 trials per signal level. All fitted psychometric functions demonstrated an r^2 of >0.5.

3. Results and discussion

Threshold, slope, criterion, and r^2 estimates are provided in Table 1. Thresholds for the noise masker were uniform across subjects and temporal conditions, ranging from 33.3 to 35.5 dB SPL for the temporally uncertain condition and from 30.0 to 34.6 dB SPL for the temporally defined condition. Thresholds were more variable for the random two-tone masker, ranging from 17.2 to 47.5 dB SPL for the temporally uncertain condition and from 5.2 to 32.2 dB SPL for the temporally defined condition.

Figure 1 presents the amount of masking release associated with defining the listening interval (temporally uncertain minus temporally defined threshold) for each subject and for the average across subjects. Data for the noise and random two-tone masker are shown by open circles and filled squares, respectively. The average effect of signal-temporal uncertainty on detection for a 1-kHz tone in the presence of a continuous sequence of noise bursts was 2 dB. This result is similar to previously reported estimates (e.g., Watson and Nichols, 1976; Green and Weber, 1980), confirming that signal-temporal uncertainty has a minimal effect on pure-tone detection in noise. In contrast to the results for the noise masker, the average effect of signal-temporal uncertainty for the random two-tone masker was 9 dB.

A one-way repeated-measures analysis of variance (ANOVA) on threshold was performed for each masker type. Thresholds were significantly higher for the temporally uncertain versus the temporally defined condition for the random two-tone masker [$F(1,3)=12.9$; $p=0.04$], but were not significantly different for the noise masker [$F(1,3)=2.4$; $p=0.2$]. These findings are consistent with the hypothesis that the effect of signal-temporal uncertainty is greater for stimuli that generate substantial informational masking than for stimuli that generate little or no informational masking.

As shown in Table 1, subjects' psychometric functions tended to have a shallower slope for the random two-tone compared to the noise conditions. This observation is consistent with previous work (e.g., Lutfi *et al.*, 2003). Note, however, that no systematic differences in slope were observed across temporally uncertain and temporally defined conditions for either masker in the current study.

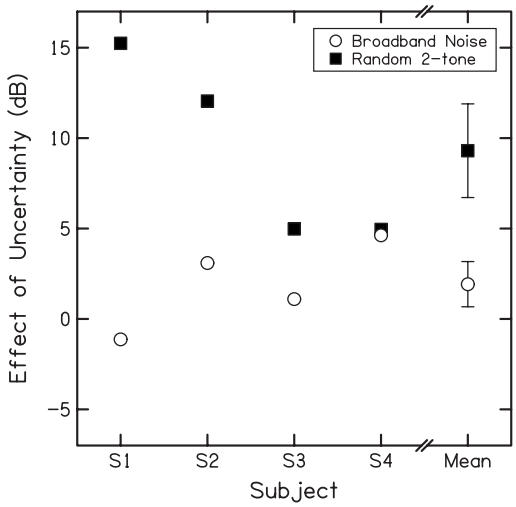


Fig. 1. The release from masking produced by reducing signal-temporal uncertainty (temporally uncertain minus temporally defined threshold) is plotted for individual subjects and for the mean across subjects. Error bars indicate ±1 standard error of the mean across subjects. Masking release for the broadband noise masker is shown by open circles. Masking release for the random two-tone masker is shown by filled squares.

As previously reported (e.g., Neff and Dethlefs, 1995), individual differences in performance were large for both random two-tone masker conditions. Estimates of the standard error for thresholds averaged across subjects for the random two-tone masker were 7.0 and 6.0 dB for temporally uncertain and temporally defined conditions, respectively. In sharp contrast, estimates of the standard error were 0.4 and 1.0 dB for temporally uncertain and temporally defined conditions, respectively. Moreover, psychometric functions fitted to the noise data accounted for a greater proportion of variance than psychometric functions fitted to the random two-tone data. As shown in Table 1, values of r^2 for the noise masker ranged from 0.83 to 0.94 when the listening interval was uncertain and from 0.90 to 0.98 when the listening interval was defined. In contrast, values of r^2 for the random two-tone masker ranged from 0.51 to 0.77 when the listening interval was uncertain and from 0.54 to 0.86 when the listening interval was defined. Thus, both between-subjects and within-subjects variability appear to be greater for the random two-tone masker relative to the noise masker.

The results of Best *et al.* (2007) support the idea that the effect of signal-temporal uncertainty is greater for conditions expected to produce substantial informational masking. Best *et al.* found that the benefit of knowing when to listen was larger for identifying birdsong than for recalling digits embedded in either forward or reverse speech. The authors suggested that this discrepancy may reflect differences in target-masker similarity across conditions. The current results for the random two-tone masker are in agreement with this suggestion. Both the target and masker stimuli were comprised of pure tones, presumably resulting in high target-masker similarity. A related explanation is that the current results reflect a combined effect of signal-temporal and masker-spectral uncertainty. Future work examining performance across multi-tonal maskers with random and fixed spectra might provide insight into the contributions of both uncertainty and similarity for these stimuli and conditions.

This study was based on that of Leibold and Werner (2003), which found an average effect of signal-temporal uncertainty of 16 dB using similar stimuli, but used a single-interval, adaptive procedure. Consistent with the earlier study, the current results indicate a larger effect of signal-temporal uncertainty on detection for a 1-kHz tone in the presence of a random two-tone masker than has been observed for tone-in-noise detection (e.g., Green and Weber, 1980). However, the average effect of signal-temporal uncertainty for the random two-tone masker was reduced in the current study (9 dB) compared to the previous work (16 dB). This discrepancy

might reflect individual differences. Note that a 15-dB effect was observed for one of the four subjects in the current study, similar to the average effect reported by Leibold and Werner. Alternatively, differences in response bias across temporally uncertain and temporally defined conditions may have contributed to the larger effect observed by Leibold and Werner. Researchers have noted that subjects may adopt a conservative bias for single-interval tasks (e.g., Marshall and Jesteadt, 1986). The influence of this conservative bias on pure-tone thresholds in quiet, however, appears to be small (e.g., Swets *et al.*, 1961; Marshall and Jesteadt, 1986). To examine whether the effect of response bias was larger for the current multi-tonal stimuli, individual estimates of criterion were calculated. As shown in Table 1, subjects appeared to be more conservative listeners in the current study compared to Leibold and Werner (2003). For the random two-tone masker, the criterion estimates ranged from 0.1 to 0.8 for the temporally uncertain condition and from 0.2 to 0.8 for the temporally defined condition. In contrast, Leibold and Werner observed a range of −1.4 to 0.8 for the adaptive temporally uncertain condition and −0.5 to 0.3 for the adaptive temporally defined condition. Further investigation is required to determine whether the apparent increase in criterion effects extends to other stimuli and procedures used to study informational masking.

In summary, this study examined whether the effect of signal-temporal uncertainty was larger for a masker expected to produce substantial informational masking than for a masker expected to produce minimal informational masking. The results confirm those of previous studies (e.g., Green and Weber, 1980), indicating a 2-dB effect of signal-temporal uncertainty for tone-in-noise detection. As predicted, the effect of signal-temporal uncertainty was significantly greater for the random two-tone masker, with effects ranging from 5 to 15 dB across subjects. These results are consistent with the hypothesis that the effect of signal-temporal uncertainty is higher for conditions in which baseline informational masking is high.

Acknowledgments

This work was supported by NIH Grant No. R03 DC008389. We are indebted to Lynne Werner and Donna Neff for their contributions to earlier work. Thanks are also extended to Emily Buss and John Grose for their helpful comments on a previous version.

References and links

Best, V., Ozmeral, E. J., and Shinn-Cunningham, B. G. (**2007**). "Visually-guided attention enhances target identification in a complex auditory scene," J. Assoc. Res. Otolaryngol. **8**, 294–304.

Chang, P., and Viemeister, N. F. (**1991**). "Temporal windows for signals presented at uncertain times," J. Acoust. Soc. Am. **90**, 2248.

Dai, H. (**1995**). "On measuring psychometric functions: A comparison between the adaptive up-down and constant-stimulus method," J. Acoust. Soc. Am. **98**, 3135–3139.

Durlach, N. I. (**2006**). "Auditory masking: Need for improved conceptual structure," J. Acoust. Soc. Am. **120**, 1787–1790.

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

Egan, J. P., Greenberg, G. Z., and Schulman, A. I. (**1961**). "Interval of time uncertainty in auditory detection," J. Acoust. Soc. Am. **33**, 771–778.

Green, D. M., and Weber, D. L. (**1980**). "Detection of temporally uncertain signals," J. Acoust. Soc. Am. **67**, 1304–1311.

Jones, M. R., and Yee, W. (**1993**). "Attending to auditory events: the role of temporal organization," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by S. McAdams and E. Bigand (Oxford U. P., Oxford).

Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colbun, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

Leibold, L. J., and Werner, L. A. (**2003**). "Infants' detection in the presence of masker uncertainty," J. Acoust. Soc. Am. **113**, 2208.

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.

Lutfi, R. A., Kistler, D. J., Callahan, M. R., and Wightman, F. L. (**2003**). "Psychometric functions for informational masking," J. Acoust. Soc. Am. **114**, 3273–3282.

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.

Neff, D. L., and Dethlefs, T. M. (**1995**). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," J. Acoust. Soc. Am. **98**, 135–134.
Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.
Swets, J. A., Tanner, W. P., and Birdsall, T. G. (**1961**). "Decision processes in perception," Psychol. Rev. **68**, 301–340.
Watson, C. S., and Nichols, T. J. (**1976**). "Detectability of auditory signals presented without defined observation intervals," J. Acoust. Soc. Am. **59**, 655–668.

Linear frequency modulation reverberation suppression using difference of singular values

Hyung Soo Lee and Kyun Kyung Lee
School of Electrical Engineering and Computer Science, Kyungpook National University, Daegu 701-702, Republic of Korea
balbari@ee.knu.ac.kr, kklee@ee.knu.ac.kr

Abstract: This letter proposes a reverberation suppression method to improve the target detection probability in reverberation environment induced by a linear frequency modulation signal. Using the singular value decomposition and an assumption of local spatial stationarity, the proposed algorithm suppresses reverberation utilizing the difference of singular values between the reference beam and the current beam of interest. Several experiments using real oceanic reverberation data justify the validity of the proposed algorithm.

© 2008 Acoustical Society of America
PACS numbers: 43.60.Gk, 43.30.Vh, 43.60.Hj [JC]
Date Received: July 15, 2008 **Date Accepted:** August 19, 2008

1. Introduction

In active sonar, the target detection is often limited by the presence of reverberation. The reverberation is caused mainly by multiple reflections, diffusions, or diffractions of the transmitted signal by the surface and the bottom of ocean interfaces. Since the reverberation is strongly correlated with the transmitted signal, the conventional detection method such as matched filtering (MF) is inefficient. In order to improve the performance of target detection, the principal component inverse (PCI) algorithm, which considers reverberation as a sum of undesirable returns reflected from many point scatterers, has been suggested to estimate and suppress reverberation before applying the conventional MF.[1] The PCI algorithm utilizes a low rank approximation of reverberation subspace via singular value decomposition (SVD) and then uses target echo power as a metric to estimate reverberation subspace rank. However, the target echo power, which is used as a metric, is seldom available in real applications.

Accordingly, this letter proposes a method to suppress linear frequency modulation (LFM) reverberation, which subtracts the singular values of the reference beam that is assumed to contain the reverberation only from those of the current beam of interest, which is assumed to contain the reverberation and target echo. As the proposed algorithm uses the left or right beam adjacent to the current beam of interest as the reference beam, we do not need a priori knowledge about the target echo power. For the validation, the proposed algorithm is applied to real oceanic reverberation data, which are acquired using the hull-mounted sonar (HMS).

2. Proposed algorithm

The PCI and the proposed algorithm, which are particular cases of subspace methods, are used to separate reverberation and the target echo by estimating the reverberation subspace. These algorithms model reverberation as a sum of echoes issued from the transmitted signal and assume that reverberation power is much greater than the target echo power to separate the reverberation subspace and the target subspace. Also, these algorithms assume that the rank of the matrix, which spans the reverberation subspace, is low to obtain good separation.[1] The subspace methods require a generating matrix. Kumaresan *et al.*[2] proposed a matrix built from the received signal $\mathbf{x}$ by cutting it into blocks $\mathbf{X}_i$. This matrix, noted $\mathbf{Y}_i$ on block i, is called the forward matrix

 © 2008 Acoustical Society of America

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{X}_i(p) & \mathbf{X}_i(p-1) & \cdots & \mathbf{X}_i(1) \\ \mathbf{X}_i(p+1) & \mathbf{X}_i(p) & \cdots & \mathbf{X}_i(2) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_i(L) & \mathbf{X}_i(L-1) & \cdots & \mathbf{X}_i(L-p+1) \end{bmatrix}, \tag{1}$$

where L is the block length and p is chosen close to $L/2$. $\mathbf{X}_i(k)$ is the kth scalar element of the sample of the ith block of $\mathbf{x}$. The forward matrix $\mathbf{Y}_i$ on block i is decomposed into two matrices $\mathbf{Y}_i^m$, which span the reverberation subspace and $\mathbf{Y}_i^o$, which spans the target plus white noise subspace as follows:

$$\mathbf{Y}_i = \mathbf{Y}_i^m + \mathbf{Y}_i^o, \tag{2}$$

where m is the rank of reverberation subspace and o is the rank of target plus white noise subspace. The rank of $\mathbf{Y}_i$ is R, which is $m+o$. As reverberation is assumed to be stronger than the target echo, $\mathbf{Y}_i^m$ is built with the largest singular values of $\mathbf{Y}_i$. From the SVD of $\mathbf{Y}_i$ and the Eckart and Young theorem,[3]

$$\mathbf{Y}_i = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{V}_i^H = [\mathbf{U}_i^m \quad \mathbf{U}_i^o] \begin{bmatrix} \mathbf{\Sigma}_i^m & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_i^o \end{bmatrix} [\mathbf{V}_i^m \quad \mathbf{V}_i^o]^H, \tag{3}$$

where $\mathbf{U}_i$ and $\mathbf{V}_i$ are the left and right singular vector maries of $\mathbf{Y}_i$, and $\mathbf{\Sigma}_i$ is a diagonal matrix that contains the decreasing singular values of $\mathbf{Y}_i$,$\{\sigma_j\}(\sigma_1 > \sigma_2 > \cdots > \sigma_R)$.

In the case of PCI algorithm, rank estimation is required to separate reverberation subspace and target subspace. The rank is estimated by computing the sum of the singular values of $\mathbf{Y}_i$ and comparing it to a threshold P, which is linked to a priori knowledge about the target echo power. If the sum of the singular values of $\mathbf{Y}_i$ is not greater than P, the PCI algorithm is not applied to block $\mathbf{X}_i$. If there exists an index M such as $\Sigma_{q=0}^{M} \sigma_{R-q}^2 > P$, then the PCI algorithm is applied to block $\mathbf{X}_i$ and the reverberation subspace rank m is equal to $R-M+1$. A vector $\mathbf{X}_i^m$ is collected from $\mathbf{Y}_i^m$ in the reverse process of generating the forward matrix (1), and then the vector $\mathbf{x}^m$ is made for all the data blocks. Finally, the MF is applied to the vector $\mathbf{x} - \mathbf{x}^m$. However, a priori knowledge about target echo power, which is used as a metric to estimate reverberation subspace rank, is seldom available in real applications.

To overcome this problem, this letter proposes a method to suppress reverberation by subtracting the singular values of the reference beam from those of the current beam of interest. As the proposed algorithm uses the left or right beam adjacent to the current beam of interest as the reference beam to estimate reverberation subspace, the proposed algorithm does not have to separate reverberation subspace and target subspace in the current beam of interest. As a result, the proposed algorithm does not need a priori knowledge about target echo power, which is necessary to estimate the rank to separate reverberation subspace and target subspace in the PCI algorithm. To suppress reverberation efficiently using adjacent beam, it is assumed that the signal is spatially stationary in the neighboring blocks, that is, reverberation energy does not change significantly between the adjacent beams.[4] The stationarity is validated in the next section.

In the proposed algorithm, we define the current beam signal of interest, which is assumed to contain the reverberation and target echo as the received signal $\mathbf{x}$ in the PCI algorithm. Moreover, we suppose that the reference beam signal contains the reverberation only. $\mathbf{Y}_{R_i}$, the forward matrix built from the reference beam, is decomposed into two matrices $\mathbf{Y}_{R_i}^n$, which span the reverberation subspace and $\mathbf{Y}_{R_i}^w$, which spans the white noise subspace as follows:

$$\mathbf{Y}_{R_i} = \mathbf{Y}_{R_i}^n + \mathbf{Y}_{R_i}^w, \tag{4}$$

where n is the rank of reverberation subspace and w is the rank of white noise subspace in the reference beam. The rank of $\mathbf{Y}_{R_i}$ is R, which is equal to the rank of current beam of interest on block i. From the SVD of $\mathbf{Y}_{R_i}$,

$$\mathbf{Y}_{R_i} = \mathbf{U}_{R_i}\mathbf{\Sigma}_{R_i}\mathbf{V}_{R_i}^H = [\mathbf{U}_{R_i}^n \quad \mathbf{U}_{R_i}^w]\begin{bmatrix}\mathbf{\Sigma}_{R_i}^n & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_{R_i}^w\end{bmatrix}[\mathbf{V}_{R_i}^n \quad \mathbf{V}_{R_i}^w]^H. \tag{5}$$

For a LFM signal with duration T and bandwidth B, the intrinsic bandwidth is $\alpha=(B/T)^{1/2}$.[5] This parameter corresponds to the smallest bandwidth, which allows separation of two LFM echoes. We recall that $\mathbf{Y}_i^m$ in Eq. (2) and $\mathbf{Y}_{R_i}^n$ in Eq. (4) represent the matrix that spans the reverberation subspace in the current and the reference beam, respectively. As the adjacent beams in temporally synchronized data blocks have similar reverberation energy by spatial local stationary assumption, $\mathbf{Y}_i^m$ and $\mathbf{Y}_{R_i}^n$ have the similar singular values. Also $\mathbf{Y}_i^m$ and $\mathbf{Y}_{R_i}^n$ have the same rank as G. Ginolhac and Jourdain[1] have mentioned that a rank of one data block is the number of times the intrinsic bandwidth is contained in the data block. That is, $\mathbf{\Sigma}_i^m$ is similar to $\mathbf{\Sigma}_{R_i}^n$ and m is equal to n. Subtracting $\mathbf{\Sigma}_{R_i}$ of Eq. (5) from $\mathbf{\Sigma}_i$ of Eq. (3), the singular values $\mathbf{\Sigma}_i^m$ corresponding to reverberation subspace are rejected and the singular values $\mathbf{\Sigma}_i^o$ corresponding to the white noise are also rejected. As a result, a diagonal matrix $\tilde{\mathbf{\Sigma}}_i=\mathbf{\Sigma}_i-\mathbf{\Sigma}_{R_i}$, which contains singular values corresponding to target only, is estimated. If a negative singular value difference exists in this process, proposed algorithm replaces this negative singular value difference with zero. Inserting $\tilde{\mathbf{\Sigma}}_i$ instead of $\mathbf{\Sigma}_i$ in (3), the forward matrix $\tilde{\mathbf{Y}}_i$ is estimated as follows:

$$\tilde{\mathbf{Y}}_i = \mathbf{U}_i\tilde{\mathbf{\Sigma}}_i\mathbf{V}_i^H. \tag{6}$$

A vector $\tilde{\mathbf{X}}_i$ is then collected from $\tilde{\mathbf{Y}}_i$ and the estimated current beam signal $\tilde{\mathbf{x}}$ is made for all the data blocks. Finally, the MF is applied to $\tilde{\mathbf{x}}$ to detect the target signal.

3. Experiment and results

For comparison, simulated targets 1 and 2, located at 1500 in port of own ship and 2250 m in front of the own ship, respectively, were injected into real oceanic reverberation data acquired in the littoral water of South Korea using HMS. To analyze the performance according to the target Doppler, target 1 is set to 0 kn, while target 2 has 4 kn velocity. The transmitted signal is a LFM signal with 400 Hz bandwidth, 6100 Hz center frequency, 50 ms pulse duration, and received by a cylindrical array of 64 sensors with 31 250 Hz sampling frequency. For stochastic analysis, we utilize 50 reverberation data sets, which lasted about 16 min. The sensor signals are shifted to the baseband and downsampled by a factor of 16. A conventional delay-sum beamformer is used with 64 beams. The beam number is named from 1 to 64 turning clockwise from the stern.

In this experiment, a block length of 48 samples (equal to 1/2 of the transmitted pulse duration) was selected for the PCI and the proposed algorithm. As the intrinsic bandwidth α is about 90 Hz, the quantity corresponding to the rank of one data block $B/(2\alpha) \cong 2.3$. Because this result satisfies the hypothesis of small rank (the matrix size is 25×24), the choice of L is reasonable. In the case of PCI algorithm, the threshold P is properly selected considering the power of the simulated target.

Figure 1 shows the normalized singular value distance between two beams in real oceanic reverberation data without a target. The normalized singular value distance on block i is defined as $d_i=1/(m\sigma_{\max})\Sigma_{h=1}^m|\sigma_h^{C_i}-\sigma_h^{R_i}|$, where $\sigma_{\max}$ is the largest singular value in the entire beams, $\sigma_h^{C_i}$ and $\sigma_h^{R_i}$ are the hth singular value, of the current and reference beams on block i, respectively, and m is the rank of reverberation subspace. In this simulation, we choose $m=3$ as the quantity corresponding to the rank of one data block $B/(2\alpha)\cong2.3$. The evaluation is re-

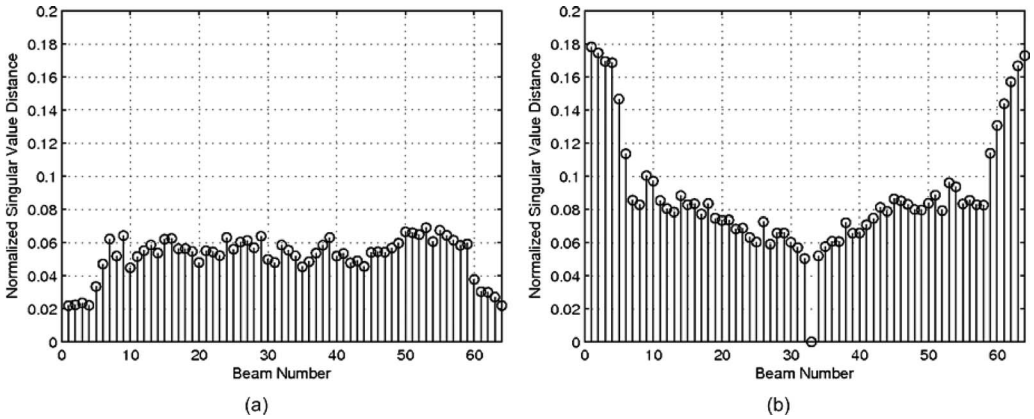


Fig. 1. Average singular value distance between two beams. (a) Average singular value distance between two successive beams. (b) Average singular value distance between beam 33 and the other beams.

peated and averaged for 50 reverberation data sets. Figure 1(a) is the normalized singular value distance between two successive beams. We select the kth beam as a current beam and the $(k-1)$th beam as a reference beam. The distance stays roughly constant. This means that reverberation energy between one beam and the next does not change very much. We plot in Fig. 1(b) the normalized singular value distance between the current beam and the reference beams. We select beam 33 as a current beam, and the other beams are selected as a reference beam. We observe that the distance is getting larger as the reference beam moves away for the current beam, while the distance of the second next beams (31 or 35) stays low. In light of these results, we will choose the second next beam of the current beam as a reference beam because there is a possibility that the target signal is smeared to the next beam.

Figure 2 shows target detection results using each method for the entire beam signal in the target injected real oceanic reverberation data. Target 1 is detected in beam number 16, which corresponds to the port of own ship, and target 2 is detected in beam number 33, which corresponds to the front of own ship. Figure 2(a) shows the MF output, and the target is difficult to find due to many false-alarm peaks. Figure 2(b) shows the MF output after the PCI algorithm. Each target peak is now visible but many false-alarm peaks still remain. Figure 2(c) shows the MF output after the proposed algorithm. The result reveals each target peak prominently due to suppression of the reverberation from all directions. This shows that the proposed algorithm works well for a stationary target and/or a moving target.

Finally, we present the detection performance of the proposed algorithm. By performing 50 Monte Carlo runs on 50 reverberation data sets, the experimental receiver operating characteristic (ROC) curves are plotted in Fig. 3. These curves represent detection probability versus false-alarm probability for each fixed signal-to-noise ratio. The detection criterion is as

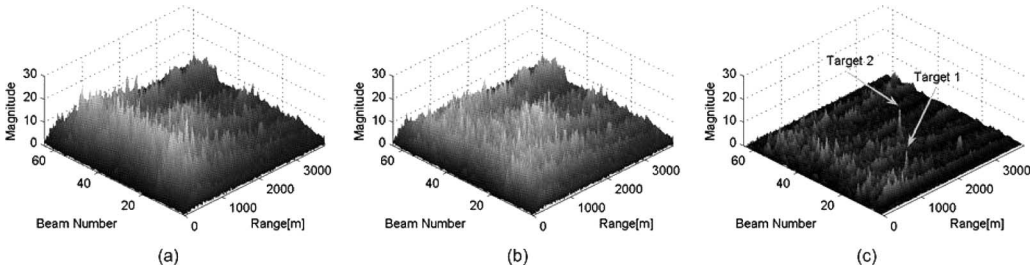


Fig. 2. Target detection results. (a) MF output. (b) MF output after the PCI algorithm. (c) MF output after the proposed algorithm.

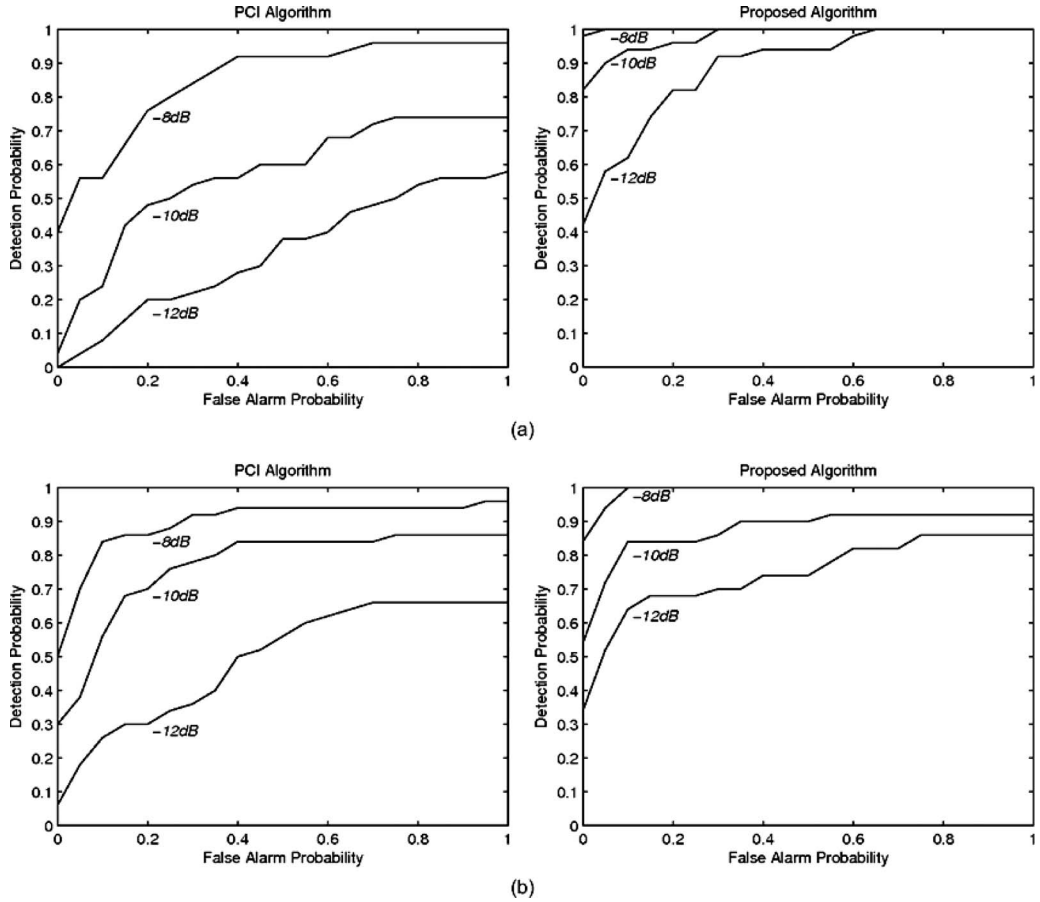


Fig. 3. Experimental ROC curves. (a) PCI (left) and proposed algorithm (right) for target 1. (b) PCI (left) and proposed algorithm (right) for target 2.

follows. When targets are absent, we evaluate the probability of false alarms on the 100 ms around the expected position of the targets as a function of detection threshold. When targets are injected, the maximum peak positions of the MF output above the detection threshold within acceptable error interval around the true positions, 50 ms in this experiment, are counted as the detection. This procedure is repeated and the results are averaged for the 50 Monte Carlo runs. Figure 3(a) represents the ROC curves for target 1 without Doppler and Fig. 3(b) represents the ROC curves for target 2 with Doppler. By comparing the left and right plots of Fig. 3, we can conclude that the proposed algorithm improves the detection performance significantly.

4. Conclusion

This letter proposes a reverberation suppression method to improve the target detection probability in reverberation environment induced by a LFM signal. By subtracting the singular values of the reference beam from those of the current beam in temporally synchronized blocks, we can suppress LFM reverberation without a prior knowledge about the target echo power. Experiments with real oceanic data verify that the proposed algorithm provides an improved reverberation suppression performance. Consequently, it is expected that the proposed algorithm can be applied for the detection of real underwater objects.

Acknowledgments

This work was supported by Defense Acquisition Program Administration and Agency for Defense Development under Contract No. UD070054AD.

References and links

[1]G. Ginolhac and G. Jourdain, ""Principal component inverse" algorithm for detection in the presence of reverberation," IEEE J. Ocean. Eng. **27**, 310–321 (2002).
[2]R. Kumaresan, I. P. Kirsteins, and D. W. Tufts, "Data adaptive signal estimation by singular value decomposition of a data matrix," Proc. IEEE **70**, 684–685 (1982).
[3]C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," Psychometrika **1**, 211–218 (1936).
[4]B. W. Choi, E. H. Bae, J. S. Kim, and K. K. Lee, "Improved prewhitening method for linear frequency modulation reverberation using dechirping transformation," J. Acoust. Soc. Am. **123**, EL21–EL25 (2008).
[5]L. R. O. Storey, "An integration of whistling atmospherics," Proc. R. Soc. London, Ser. A **246**, 113–141 (1953).

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



Ann R. Bradlow—For contributions to variation in speech intelligibility



John Fawcett—For contributions to three dimensional scattering from objects in the seabed



Keith R. Kluender—For contributions in audition and perceptual learning

Annual Reports of Technical Committees

(**See October issue for additional reports**)

Animal Bioacoustics

Animal Bioacoustics had a good year, with meetings and special sessions in New Orleans, LA in the fall, and in Paris, France in the spring. In addition, AB is sponsoring a special meeting on Animal Sound Communication, to take place in Corvallis, Oregon on August 12-15, 2008. Many thanks to David Mellinger and Sara Heimlich of Oregon State University who organized and are coordinating the meeting. This is the second in a series of Animal Sound Communication meetings that have been sponsored by Animal Bioacoustics and The Acoustical Society of America, and promises to be very exciting and packed with interesting papers.

In New Orleans, the AB Chair was unable to attend, but was temporarily replaced by past-Chair Mardi Hastings. Many thanks to Mardi. Animal Bioacoustics sponsored one special session on mechanisms of sound source localization in a wide range of animals, from fishes to humans that was organized by Michael Dent and Eduardo Mercado.

The Paris meeting was very well attended as were the Animal Bioacoustics sessions. Our committee is very grateful to the European co-Chair Michel André (Laboratori d'Aplicacions Bioacústiques, Universitat Politècnica de Catalunya (UPC)) for his work as co-Chair, as the AB TPOM rep in Paris, as co-chair of several structured sessions in Paris, and as one of the judges for best student paper awards. The winning paper was by Yossi Yovel of Eberhard-Karls-Universitat Tubingen, Germany with a paper on "The statistics of plant echoes perceived by echolocating bats." AB decided to award students presenting paper separately from those presenting posters. The winning poster was by Joseph Jackson, University of Bristol, UK. Many students participated in the Paris meeting, and judging was a task shared by many AB members. The Paris meeting included many special/structured sessions as well as general sessions. The opening ceremony included a talk by AB member Peter Tyack on "How sound from human activities affects marine mammals." The special sessions were: Anthropogenic Noise Effects on Animals (co-sponsored by ECUA and the Noise TC, and chaired by Michael Stocker and Jonathan Gordon), Animal Sonar Systems (co-sponsored with AB by Underwater, Acoustical Oceanography, Signal Processing, and ECUA, and co-chaired by James Simmons and Michel Andre), Animal Bioacoustic Censusing (co-sponsored by ECUA and co-chaired by Marie Roch and Cederic Gervaise), Auditory Brainstem Response and Behavior Correlations co-sponsored by ECUA and Psychological and Physiological Acoustics, and Co-chaired by Elizabeth Britten-Powell and Alexander Supin), Odontocete Acoustics (co-sponsored with ECUA and co-chaired by David Mellinger and Michel André), and Sound Production and Reception in Amphibious Marine Mammals (co-chaired by Jason Muslow, Ronald Schusterman, and Ian Boyd).

Plans for the Miami meeting in the fall include a special session honoring Sam Ridgway (co-chaired by Whitlow Au and Dorian Houser) and a session on the acoustics of manatees and alligators co-chaired by Ann Bowles and Jennifer Miksis-Olds. The TPOM rep will be Dorian Houser. For the Portland meeting, special sessions were suggested on noise effects on terrestrial animals (Kurt Firstrup), Marine mammal population assessment using acoustics (David Mellinger), Autonomous remote monitoring systems (Marc Lammers), and Acoustical monitoring to track biotic changes (Kurt Firstrup and Ann Bowles).

Richard Fay
Chair

Biomedical Ultrasound/Bioresponse to Vibration

Special sessions (and organizers) at the New Orleans meeting included: Topical Meeting on Tissue-response to acoustics and vibrations (Tyrone Porter), Acoustic Tomography in Tissue (Jeffrey Ketterling, Iwaki Akiyama), Stable Cavitation (Saurabh Datta and Glynn Holt), and Biomedical

 0001-4966/2008/124(5)/2647/9/$23.00 © 2008 Acoustical Society of America

Applications of Acoustic Radiation Force (Mostafa Fatemi). Charles Church was the TPOM representative.

Special sessions (and organizers) at the Paris meeting included:. High-intensity focused ultrasound (Lawrence Crum and Jean-Yves Chapelon), High-intensity focused ultrasound metrology and standards (Peter Kaczkowski and Gail ter Haar), Shock waves in medicine (Robin Cleveland and Oleg Sapozhnikov), Theoretical and computational models of ultrasonic propagation in bone (James Miller and Pascal Laugier), Ultrasonic characterization of bone (Keith Wear and Frédéric Padilla), Quantitative ultrasound methods for diagnosis and therapy (Jonathan Mamou and Geneviève Berger), Light and sound in medical imaging and therapy (Ronald Roy and Emmanuel Bossy), Transducers for medical imaging and therapy (Jeffrey Ketterling and Marc Lethiecq), Ultrasound contrast agents for therapy (Tyrone Porter and Eleanor Stride), and Biomedical Applications of Acoustic Radiation Force (Mostafa Fatemi and Mathias Fink). There was a special session Biomedical Applications of Time Reversal (Oleg Sapozhnikov and Mickael Tanter) that was part of a Topical Meeting on Time Reversal, a special session co-sponsored with PA Acoustically Activated Bubble Dynamics and Applications (Carr Everbach and Joachim Holzfuss), and a contributed session (Michael Bailey and Oleg Sapozhnikov). Neil Owen and Emmanuel Bossy were the TPOM representative who were able to fit all these sessions into the busy week.

Several members of the BBTC received awards and recognition. Tyrone Porter received the R. Bruce Lindsay Award "for contributions to ultrasound-guided drug delivery." Vera A. Khokhlova and Glynn Holt were named Fellows of the ASA. David Blackstock was awarded the Rossing Prize in Acoustics Education. Student Paper Award winners were Moire Smith of Cambridge University for "Effect of HIFU on pH responsive PEGylated Micelles," Todd Hay of The University of Texas for "Ultrasound Frequency response of bubble pulsations in tubes with arbitrary wall impedance," and Dorothée Bossis of Université Pierre et Marie Curie for "1 kHz sound stimulates nitric oxide and prostaglandin E2 production by rat mesenchymal stem cells."

A number of people should be acknowledged for their work for the BBTC and new volunteers are introduced. The associate editors for BB are Charles Church, Floyd Dunn, and Douglas Miller. Shira Broschat continues to maintain the BBTC website (http://moab.eecs.wsu.edu/~shira/asa/bubv.html). Lawrence Crum is the representative to the Medals and Awards Committee. Christy Holland replaces Carr Everbach as BB representative on the Membership Committee and Peter Kaczkowski replaces Carr Everbach on ASACOS. Paul Barbone has served as the ASA representative to the Biomedical Imaging Research Opportunities Workshop (BIROW) and will be replaced by Subha Maruvada. Michael Canney of the University of Washington became the Chair of the Student Council. Lucie Somaglino of INSERM U556 and University Claude Bernard, Lyon, France is the BBTC representative to the Student Council. Jeffrey Ketterling has been elected Chair of the BBTC.

Michael Bailey
Chair (2006-2008)

Physical Acoustics

It has been an honor for me to be your Chair and a pleasure to have the opportunity to reflect back on the past year's activities in which Physical Acoustics has been involved. It is the talent, dedication and professionalism of PA committee members (working together with ASA members across the board) on the long range planning that goes into forming the ASA activities and special sessions that blossom at future ASA meetings.

It was an exciting and productive year for the Physical Acoustics Technical Committee of the ASA. The tremendous success of the 154th Meeting (held in New Orleans from 27 Nov.–1 Dec. 2007) and the 155th Meeting, Acoustics' 08, Paris (held at the Palais des Congrès from 29 June–4 July 2008) speak volumes for your commitment to the Society.

The planning for New Orleans developed during the Hawaii and Salt Lake City Meetings. Two special sessions were organized for New Orleans: (a) Acoustic applications for hurricane and storm preparedness was organized by Jim Sabatier, Ken Gilbert and Roger Waxler. (Preston Wilson chaired the session); (b) Ultrasonic measurements and methods for condensed matter was organized by Albert Migliori and Veerle Keppens (who also served as chair).

Jim Chambers and Jim Sabatier volunteered to be the PA TPOM representatives (10–11 Aug. 2007). Contributed papers were organized into five sessions (with the chairs mentioned in parentheses): (c) General physical acoustics (Joel Mobley), (d) Bubbles and cavitation (Ron Roy), (e) Thermoacoustics (William Slaton), (f) Acoustics of soil and porous media (Wheeler Howard and Alexander Ekimov), and (g) Atmospheric acoustics (D. Keith Wilson). Thanks to Jim Chambers and Jim Sabatier for all their help.

A special thanks goes to Alfred J. Bedard, who gave an excellent Tutorial Lecture entitled, "Tutorial lecture on weather and acoustics." (Jim Chambers chaired this session along with Hot Topics in Acoustics.) One of out PATC members, William M. Carey, received the 2007 ASA Pioneers of Underwater Acoustics Medal. Congratulations Bill! David Blackstock received the Rossing Prize in Acoustics Education and gave a very insightful lecture, "Songs my students sang to me," which was a wonderful tribute to his students. Congratulations David! (Mark Hamilton as chair introduced his lecture.) Here we mention that Dimitri M. Donskoy and R. Glynn Holt became our newest Fellows in the Society. Congratulations to Dimitri and Glynn!

All said, the New Orleans Meeting was a big success for PA.

It is important to mention the launching of POMA, the Society's new online open access journal, "Proceedings of Meetings on Acoustics," which began to receive submissions on Nov. 2, 2007. Vol. 1 starts with Salt Lake City and Vol. 2 goes with the New Orleans Meeting, etc. (The general idea was conceived by the Executive Director, Charles Schmid. Anthony Atchley (the current ASA President at the time) and Allan Pierce (Editor-in-Chief) announced the details of POMA to the PATC at the Salt Lake City Meeting. Tom Matula and Murray Korman were appointed by committee to be the first PA and ED Associate Editors of POMA, respectively.

The brainstorming of special sessions in PA for the Acoustics' 08, Paris Meeting originated around the time of the Hawaii meeting. By the end of the Salt Lake Meeting the original eight special sessions had expanded to eighteen. Organizers had a narrow window until 30 June 2007 to submit various detailed information (including the European co-organizer) prior to the call for papers. Acoustics' 08 combined the 155th Meeting of the ASA, 5th Forum Acusticum of the EAA, 9e Congrès Française D'acoustique of the SFA and the 2nd ASA-EAA Joint Conference; integrating the 7th European Conference on Noise Control (euronoise), the 9th European Conference on Underwater Acoustics (ecua) and the 60th Anniversary of the SFA.

Key organizers from France who worked very closely with the PA Chair in planning special PA sessions (and who were present at the Hawaii, Salt Lake City and New Orleans meetings) were Manell Zakharia (SFA) co-chair of the Meeting Organizing Committee and Philippe Blanc-Benon (SFA) co-chair of the Technical Organizing Committee. Your Chair also worked closely with Didier Cassereau (SFA), the Information Systems Manager who helped design the web software from which the meeting was planned electronically. The PATC Chair and Walter Lauriks (EAA) were the PA Technical Program co-chairs. The Chair wishes to recognize the help he received from William Hartmann, co-chair of the Technical Organizing Committee, Elaine Moran and Charles Schmid during planning Acoustics' 08 at the Technical Council Meetings. PA also had representation from Kenneth Foote (ECUA Committee), Michael Stinson and Keith Attenborough (Euronoise Committee) and Yves Berthelot (Technical Program Organizing Committee). Peter Rogers volunteered to be our PA TPOM rep. (14–16 Feb. 2008) and worked with Walter Lauriks (EAA) to help organize the meeting in Paris. Prior to the TPOM, Jim Sabatier helped the PATC Chair read and sort all the contributed paper abstracts and grouped them into sessions with titles (a two day effort.) Walter Lauriks also played a key role in supplementing this effort.

PA had planned eighteen special ("structured") sessions in Paris. They are listed below with the ASA and European co-organizers/co-chairs in parenthesis: (1) Acoustic landmine detection—joint with SP, (James Sabatier & Keith Attenborough), (2) Acoustic probes of planetary environments, (Andi Petculescu & Martin Towner), (3) Acoustically activated bubble dynamics and applications—joint with AO, BB, (Carr Everbach & Joachim Holzfuss), (4) Ultrasonics under extreme conditions, (Albert Migliori & Frederic Decremps) (5) Infrasound, (Henry Bass & Alexis Le Pichon), (6) Sonic, ultrasonic, and megasonic cleaning, (Ralph Muehleisen & Claus-Dieter Ohl), (7) Nonlinear acoustics of unconsolidated granular media, (Paul Johnson & Vitalyi Gusev), (8) Nonlinear acoustics of consolidated materials and non destructive testing, (James TenCate & Koen Van Den Abeele), (9)

Time reversal acoustics for nonlinear imaging, (Alexander Sutin & Serge Dos Santos), (10) Nonlinear acoustics in earthquake processes and other earth processes, (Paul Johnson & Fabrice Cotton), (11) Outdoor sound propagation and uncertainties, (Keith Wilson & Michel Bérengier), (12) Phononic crystals, (Pierre Deymier & Jérôme Vasseur), (13) Photoacoustics, (Gerald Diebold & Christ Glorieux), (14) Acoustics of porous media, (Walter Lauriks & Keith Attenborough), (15) Thermoacoustics, (Steven Garrett & Philippe Blanc-Benon), (16) Ultrafast acoustics, (David Hurley & Clément Rossignol), (17) Diffraction of waves on periodical structures: Acoustic, ultrasonic and acousto-optical diffraction phenomena, (Nico Declercq & Nataliya Polikarpova), and (18) Combustion noise and thermoacoustics, (Tim Lieuwen, Maria Heckl & Rafael Piscoya).

The response to the call for papers (in PA) for Acoustics' 08 was very strong in regard to contributed ("unstructured") sessions. These papers were placed in the sessions that are listed below along with the co-chairs which volunteered their services: (The symbol p in brackets means poster session.)

(1) Diffraction of waves on periodic structures: Acoustical, ultrasonic, and acousto-optic diffraction phenomenon I, II [p], (Nico Declereq and Nataliya Polikkarpova), (2) Ducts and waveguides I [p], II, (Marc Deschamps and Oliver Dazel), (3) General topics in physical acoustics I [p], II, (Walter Lauriks and Andi Petculescu), (4) General topics in nonlinear acoustics I, II [p], (Thomas Matula and Murray Korman), (5) Mathematical and numerical methods I, II [p], (Andrew Piacsek and B.D. Zaitsev), (6) Quantum Acoustics I, II [p], (Michel De Billy and Walter Lauriks). (7) Scattering and diffraction I [p], II, (Michael L. Oelze and Jean-Marc Conoir), (8) Ultrasonics: Industrial NDT I, II[p], (Bertrand Nongaillard and Christophe Aristegui), (9) Ultrasonics: Material Characterization I, (Bart Lipkens and Nicolas Wilkie-Chancellier), II [p], (Jim A. Tencate and Koen Van Den Abeele), (10) Ultrasonics: NDT and layered systems I, II [p], Vincent Pagrieux and Claire Prada), (11) Ultrasonics: Transducers and Instruments I [p], II, (Albert Migliori and Claude Boutin), and (12) Vibration in plates I,II [p], (Daniel Royer and Oliver Poncelet).

Noise, Physical Acoustics and EURONOISE co-sponsored Aeroacoustics I, II, and III and PA co-sponsored SP sessions (a) Overview of time reversal in acoustics and (b) Biomedical applications of time-reversal.

One of the highlights of the Paris Meeting was that my European PA co-organizer, Walter Lauriks (EAA) received the Foreign Medal 2007 of the French Acoustical Society. This medal is annually presented to a renowned scientist having strong relationships with French acousticians. Congratulations Walter and thank you for all your help. Tyrone M. Porter received the ASA, R. Bruce Lindsay Award 2008 for contributions to ultrasound-guided drug delivery, and Charles E. Schmid received the ASA, Distinguished Service Citation 2008. Congratulations Tyrone and Charles!

In Paris, 25 year service awards were given to Anthony A. Atchley, Kerry Commander, Nicholas P. Chotiros, Peter H. Dahl, Robert Keolian, Timothy K. Stanton and Junru Wu. Also, 50 year service awards were given to William J. Cavanaugh, Clarence S. Clay, Jr., Walter M. Madigosky, Alan Powell and Wayne M. Wright.

All said, the Paris meeting was a huge success.

It is time to acknowledge members of PA who volunteer their services on various committees: Tom Matula currently serves on the Medals and Awards Committee. Anthony Atchley, Robert Keolian and James Sabatier, all predecessors of this position remain active in helping on this Committee. Steve Garrett serves on the Membership Committee; Phil Marston serves on the Books Committee, and Richard Raspet serves on ASACOS. Todd Hay has done an outstanding job as our Student Council Rep. His term has come to an end and a subcommittee headed by David Blackstock, Anthony Atchley, Kent Gee and Todd Hay headed a selection process for the new rep. Jon R. La Follet has been selected as the next Student Council Rep. and Kenneth B. Bader will serve as the alternate. Congratulations Jon and Ken!

Your Chair has had the good fortune to serve during the presidencies of Anthony Atchley, Gilles Daigle and most recently Mark Hamilton. PA is well represented in other areas as, Vic Sparrow is our Vice President, Gilles Daigle and Wayne Wright serve on the Executive Council, and Ron Roy serves on the Technical Council. Keith Wilson is our editor of JASA-EL.

The 2008 Physical Acoustics Summer School, PASS 2008, took place from May 25–June 01, 2008, at Sunrise Springs, La Cienega, New Mexico. The program is jointly run by NCPA (George Atkins PASS 2008 Coordinator) and Penn State (Anthony Atchley), with strong ASA ties. Meetings take place every two years.

It is important to note that the ASA is a sponsor of the 18th International Symposium on Nonlinear Acoustics (ISNA) which was held in Stockholm, July 7–10, 2008. PA members lend support as Anthony. Atchley, Mack Breazeale, Larry Crum, Mark Hamilton and Paul Johnson served both on the Program Committee and the International Organizing Committee Larry Crum and Vera Khokhlova organized a special session entitled, "Nonlinear Acoustics in Medicine." for this ISNA.

Finally, I wish to thank the Physical Acoustics community from the bottom of my heart for giving me the opportunity to serve as Chair over the past three years. It has been an honor and a privilege to serve you. It has been a humbling experience for me for you to entrust me with this duty. You are a group knitted in great friendships and I would like to thank many people for your advice and encouragement over the years. Now, I pass the crown to our new Chair, Ron Roy. Have a great time, Ron; it was a great experience for me!

Thank you, PATC members, for all your help.

MURRAY S. KORMAN
Chair 2005-2008

Structural Acoustics and Vibration

At the 153rd ASA Meeting in Salt Lake City, Utah, the Structural Acoustics and Vibration Technical Committee (SAVTC) sponsored two special sessions: 1) Launch Vehicles and Space Vehicles, organized by James Phillips; and 2) Sound Source Localization, organized by Sean Wu. The Best Student Paper Competition was conducted by James Phillips and the winners were Samuel D. M. Adams (1st Place), Richard E. Dziklinski III (2nd Place). At the SAVTC meeting, two major issues were discussed: 1) Proposed online publication of conference papers; and 2) Intra-Society Activities Committee. Allan D. Pierce gave a brief summary of the characteristics of the online publication of conference papers. Courtney Burroughs proposed the formation of the Intra-Society Activities Committee to provide a forum to exchange information on the technical and administrative activities in other professional societies involved in acoustics that may be useful to ASA. Courtney Burroughs reported that there were four nominations for Fellow from SAVTC. Student Council reported that there will be a new student chapter at the University of Hartford and a student mentoring award will be presented by the Student Council in New Orleans. Finally, revision of PACS SAVTC numbers was completed by the subcommittee chaired by Earl G. Williams and submitted to Allan D. Pierce, Editor-in-Chief of JASA.

At the 154th ASA Meeting in New Orleans, SAVTC sponsored two special sessions on: 1) Modeling of Vibration and Radiation in Complex Structures, organized by Courtney Burroughs; and 2) Ground Vibration Impact in Buildings, organized by James Phillips. James Phillips also conducted the Best Student Paper Competition and the winners were Kyungmin Baik (1st Place) and R. Benjamin Davis (2nd Place). At the SAVTC meeting, some Technical Initiatives were discussed. One of them was collection of success stories in acoustics. James Phillips has submitted a story on "A Floating Opera House for Toronto." The second issue was on POMA (Proceedings of Meetings on Acoustics). Steve Shepard reported that POMA papers should be written versions of presentations made at ASA meetings. There will be $100 publication fees for each POMA paper. The window for submission of POMA papers is typically six months before the next ASA meeting. Allan D. Pierce discussed the opportunity to get funding from NAH to create special lectures and short courses for the Institute of Acoustics, which is mainly targeting post-doctoral fellows and young researchers in other American countries, excluding U.S. and Canada, to get them involved in research activities in acoustics.

At the 155th ASA/EAA Joint Meeting in Paris, SAVTC co-sponsored one general session and eight special sessions on: 1) General Topics in Structural Acoustics and Vibrations, organized by Wolfgang Kropp and Sean F. Wu; 2) Vibration and Radiation from Complex Structural Systems, organized by David Feit and Jean-Louis Guyader; 3) Source Characterization in Structure Borne Noise Problems, organized by Evan Davis and Charles Pezerat; 4) Acoustic Imaging in Confined Space, organized by Earl G. Williams and Alexandre Garcia; 5) Efficient Boundary Element Methods, organized by Ramani Duraiswami and Lothar Gaul; 6) Fluid-Structure Interaction, organized by Noureddine Atalla, Vicente Cutanda Henriquez, and Stefan Schneider; 7) Distributed Active Noise and Vibration Control, organized by Kenneth Cunefare and Manuel Collet; 8) Ground Vehicle Noise

and Vibration, organized by Donald B. Bliss and Paul De Vos; and 9) Active Noise Control: New Strategies and Innovative Concepts, organized by Alain Berry and Marie-Annick Galland. Wolfgang Kropp and Sean F. Wu conducted the Best Student Paper Competitions and the winners were: Christensen Rene (1st Place), Moreau Danielle (Tie 2nd Place) and Ege Kerem (Tie 2nd Place). Mauro Pierucci gave a summary of the guidelines of the ASA Medals and Awards Committee. Allan D. Pierce reported that the revision of PACS SVATC numbers submitted by the subcommittee chaired by Earl Williams looked good, and wanted to go through all PACS and move all papers on solids and vibration, elastic waves, and porous media to SAV. Jerry Ginsberg reported that Rona Ginsberg is currently organizing 101 Careers in Acoustics. Those who have stories or helpful information in this regard should contact Rona Ginsberg.

Sean Wu
Chair

Underwater Acoustics

The New Orleans meeting in November 2007 was a very successful one for the Underwater Acoustics Technical Committee (UWTC), with 10 sessions spread over five days, including special sessions organized by Eric Thorsos and John Perkins on Underwater Reverberation Measurements and Modeling, and a session on the Design of Distributed Surveillance and Oceanographic Monitoring Systems, chaired by Art Baggeroer and Juan Arvelo. Many thanks to our New Orleans TPOM representatives Guy Norton, William Sanders and Richard Keiffer (along with Technical Program Chairs George and Juliette Ioup) for organizing all the sessions together with the special session organizers. The best student papers in underwater acoustics were "Scattering analysis accelerated by a 3-D multilevel non-uniform grid field evaluation algorithm" by Yaniv Brick of Tel Aviv University and "Effect of resolution on synthetic aperture sonar image statistics" by Shawn Johnson of Penn State University. New fellows named at the New Orleans meeting included Ron Wagstaff, Steve Finette and Tony Lyons. Congratulations to all.

The New Orleans meeting was followed with the 155th meeting of the Acoustical Society of America, joint with the 9th European Conference on Underwater Acoustics, the 5th Forum Acusticum of the European Acoustics Association, the 9th Congres Francais d'Acoustique of the Societe Francaise Acoustique, and the 7th Euronoise, in Paris, France. It was an exciting meeting with an exhausting schedule. There were 23 sessions directly sponsored by the UWTC alone, plus another seven in Acoustical Oceanography. The UW technical committee would like to extend its warmest thanks for a very successful meeting to the co-organizers Bertrand Dubus of the SFA, Michael Vorlander of the EAA, Bill Yost of the ASA and Manell Zakharia of the SFA, as well as to Technical program Organizing Committee of Philippe Blanc-Benon of the SFA, Dick Botteldooren of the EAA, Bill Hartmann of the ASA and Pascal Laugier of the SFA. We would also like to thank the organizers of the ECUA section of the program David Bradley of the ASA and Yann Stephan of the SFA. Your TC chair had the pleasure of working closely with Yann during the frenetic but highly efficient TPOM, which went off without a hitch thanks to the tremendous advance work by Manell, Philippe and Didier Cassereau, the designer of the conference organization software, and many others.

The TC would like to thank the 23 "pairs" of special session organizers for the Paris meeting, whose work went a very long way towards making the meeting especially memorable and productive. Since this report is part of an archival document I will list them here: 1) Peter Dahl and Mike Ainslie for Seabed and Sea Surface Interaction Measurements and Modeling, 2) Henrik Schmidt and Finn Jensen who organized Fifty Years of Progress in Sonar Acoustic Research: the Role of NURC/SACLANTCEN, 3) Paul Hines and Paul White who organized Auralization of Sonar Signals, 4-5) Ken Foote and Stephen Robinson who organized Sonar System and Transducer Calibration Methodology as well as Determination of Acoustic Properties of Materials for Sonar Applications, 6) Kevin Smith and Jean-Pierre Hermand for Acoustic Vector Fields and Sensor Processing, 7) Dajun Tang, Paul Hines and Peter Nielsen for Impact of Environmental Variability on Mid-Frequency Sonar Performance, 8) Nick Chotiros and Jean-Pierre Sessarego who organized Geoacoustic Sediment Modeling, 9) Jim Lynch and Thomas Folegot who organized Impact of Internal Waves on Shallow Water Propagation, 10) Ray Soukup and J. P. Sessarego for Tank Experiments, 11) Marcia Isakson and Thomas Folegot who organized High Frequency Variability, 12) Benjamin Dzikowicz and Eugeniusz Kozaczka who chaired Nonlinear Acoustic Methods in Searching for Buried Objects, 13) Andrew Seybert, Martin Ochmann, and Bodo Nolte for Numerical Methods in Underwater Scattering, 14) Heechun Song and Sergio Jesus who organized Broadband Underwater Communications, 15) Tony Lyons, Philippe Blondel and Nick Pace who organized High Frequency Scattering, 16) Peter Gough and Roy Hansen for Low-Frequency and High-Frequency Synthetic Aperture Sonar, 17) Henrik Schmidt and Stephane Jespers who organized Sensor Coalition, 18) Jim Candy, Hugh Griffiths, and Rene Garello who co-chaired Synthetic Aperture Sonar and Radar Convergences, 19) David Calvo and Michael Taroudakis who co-chaired Sound Propagation in 3D environments, 20) Eric Thorsos and Mario Zampolli who organized Scattering from Objects Near Boundaries, 21) Gerald Dobeck, Marc Pinto and Yvan Petillot for Automatic Target Recognition, Sensors and Algorithms, 22) Eric Maillard and Benoit Zerr who organized Acoustic Data Fusion, and 23) Georges Bienvenu and Ron Wagstaff who co-chaired Noise Suppression, Robust Direction of Arrival, and Target Strength Estimation. One can see by the breadth and depth of the special sessions that underwater acoustics continues to be the home to a wide variety of cutting edge research areas and that the field has attracted and kept a large number of top researchers in both Europe and in North America.

New fellows named at the Paris meeting included Subramaniam Rajan and John Osler. Results are also in on the student paper award winners. I would like to announce that the winner is Zachary J. Waters of Boston University with his talk "Laboratory investigations of the detection and characterization of buried targets by iterative, single channel time reversal," co-authored with Benjamin Dzikowicz, R. Glynn Holt, and Ronald Roy, and Jon La Follett of Washington State University with his talk "Elastic and interfacial contributions to SAS images of tilted metal cylinders: Laboratory experiments," co-authored with Kyungmin Baik and Philip Marston. We had a tremendous student paper competition at Paris, with 32 contestants and 12 judges working overtime to evaluate them. Congratulations to the winners and a heart felt thanks to the judges!

It is with great sadness that I must conclude by reporting the passing of two of UW's most active members, Ralph Goodman, former Director of SACLANTEN, NORDA, and the National Center for Physical Acoustics at the University of Southern Mississippi, and Stanley Flatté, Professor Emeritus of Physics at the University of California, Santa Cruz. Their absence will be keenly felt by their many collaborators, friends and colleagues in the Underwater Acoustics Technical Committee.

Kevn LePage
Chair

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.

	2008
10–14 Nov	156th Meeting of the Acoustical Society of America, Miami, FL [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].
12–15 Aug	2nd Symposium on Acoustic Communication by Animals, Corvallis, OR [http://oregonstate.edu/conferences/animalcommunication2008/index.html]
8–9 Sep	1st International Symposium on Audible Acoustics in Medicine and Physiology, West Lafayette, IN [https://engineering.purdue.edu/Acoustics/]
	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].

26–30 Oct	158th Meeting of the Acoustical Society of America, San Antonio, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

2010

19–23 May	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 Nov	2nd Iberoamerican Conference on Acoustics (Joint Meeting of the Acoustical Society of America, Mexican Institute of Acoustics, and Iberoamerican Federation on Acoustics), Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

Cumulative Indexes to the Journal of the Acoustical Society of America

Ordering information: Orders must be paid by check or money order in U.S. funds drawn on a U.S. bank or by Mastercard, Visa, or American Express credit cards. Send orders to Circulation and Fulfillment Division, American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2270. Non-U.S. orders add $11 per index.

Some indexes are out of print as noted below.

Volumes 1-10, **1929-1938**: JASA, and Contemporary Literature, 1937-1939. Classified by subject and indexed by author. Pp. 131. Price: ASA members $5; Nonmembers $10

Volumes 11-20, **1939-1948**: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 395. Out of Print

Volumes 21-30, **1949-1958**: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20; Nonmembers $75

Volumes 31-35, **1959-1963**: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20; Nonmembers $90

Volumes 36-44, **1964-1968**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 485. Out of Print

Volumes 36-44, **1964-1968**:Contemporary Literature. Classified by subject and indexed by author. Pp. 1060. Out of Print

Volumes 45-54, **1969-1973**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound)

Volumes 55-64, **1974-1978**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound)

Volumes 65-74, **1979-1983**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25 (paperbound); Nonmembers $75 (clothbound)

Volumes 75-84, **1984-1988**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound)

Volumes 85-94, **1989-1993**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound)

Volumes 95-104, **1994-1998**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 632, Price: ASA members $40 (paperbound); Nonmembers $90 (clothbound)

Volumes 105-114, **1999-2003**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp.616, Price: ASA members $50; Nonmembers $90 (paperbound)

REVISION LIST

Associates

Ahn, Mi Jeoung, MEDISON Co. Ltd., R&D, Discusser & Medison B/D, 1003, Daechi-dong, Gangnam-gu, Seoul, Korea

Argus, Jeffrey, DuPont, Building Innovations, Chestnut Run Plaza 721-1306, Wilmington, DE 19880

Baird, Andrew, 47 Coombe Valley Rd, Weymouth, DT3 6NL, UK

Balas, Anna, Grochowe Laki 2/36, 61-752 Poznan, Poland

Balasubramanyam, Chandrasekhar, #1, 3rd Main, 3rd Block, Goraguntepalya, Bangalore-Karnataka 560022, India

Baum, Robert, Menlo Scientific, Ltd., 340 Channing Way, #347, San Rafael, CA 94903

Beamer, Sharon, ASHA, Audiology Dept, 2200 Research Blvd #305, Rockville, MD 22850

Benki, Jose R., Communicative Sciences and Disorders, Michigan State Univ., 101 Oyer Bldg., East Lansing, MI 48824

Berraki, Madjid, Marine Sources Division, SERCEL, Z. I. Toulon Est, 150 Rue Pasteur, La Farlede 83210, France

Blackwell, Ezra, 3675 Barnard Dr., Apt. 124, Oceanside, CA 92056

Boeckx, Laurens, Huntsman Europe BVBA, Core science, Everslaan 45, BE-3078, BE-3078 Everberg, Belgium

Borja, Eric, 650 W Wayman St., 403, Chicago, IL 60661

Brown, Randy, Soundproof Windows, 4673 AirCtr. Cir., Reno, NV 89502

Cedolin, Leonardo, Rosen Goldberg Der & Lewitz, Acoustical Consultants, 1100 Larkspur Landing Cir., Suite 354, Larkspur, CA 94939

Cerda, Pedro, C/Ballester 61-63, 4o1a esc.A, 08023 Barcelona, Spain

Chowning, John, 4235 McKellar Ln., Palo Alto, CA 94306

De Jong, Christ, Fahrenheitstraat 405, 2561 DZ The Hague, The Netherlands

Donoher, Gary, Analysis, Design & Diagnostics, 317 West Forsyth St., Jacksonville, FL 32202-4307

Dudzinski, Kathleen, Geo-Marine, Inc., Marine Sciences, P.O. Box 711, Old Mystic, CT 06372-0711

El-Khateeb, Sorour, Kinetics Middle East LLC, Engineering Depts., P.O. Box 37670, Dubai, U Arab Emirates

El-Aassar, Ahmed A, Environmental Acoustics Inc., 5801 Grayson Rd., Unit 1, Harrisburg, PA 17111

Elnoshokaty, Ibrahim, Enoshmink Technology, technical, 8 Rashdan St. Elmesaha Sq., Dokki, Egypt

Erdogan, Metin, ASELSAN A.S., Istanbul Naval Systems, Mehmet Akif Ersoy Mah., 16. cadde No: 16 Yenimahalle, 06370 Ankara, Turkey

Fleming, Richard, Defence Research and Development Canada-Atlantic, Technology, Dept., 9 Grove St, Dartmouth Nova Scotia B2Y 3Z7, Canada

Fowler, Kevin, 1810 Broadway, Apt.. 21, Oceanside, CA 92054

Guillaume, Mathieu, Hofangerstr. 45a, 81735 Muenchen, Germany

Guillon, Laurent, Ecole navale, IRENav, BP 600, 29240 Brest Armies, France

Halvorsen, Michele, 9314 Cherry Hill Rd., Apt. 502, College Park, MD 20740

Handegard, Nils Olav, Observation Methodology, PO Box 1870 Nordnes, Bergen 5817 Norway

Hix, Penny, 1435 Bayville St., Apt.. 8, Norfolk, VA 23503

Hommedal, Inge, Kilde Akustikk AS, Tvildesvegen 16D, NO-5700 Voss, Norway

Isei-Jaakkola, Toshiko, Chubu Univ., Dept. of English Language, 1200 Matsumoto-cho, Kasugai-City, 487-8501, Japan

Itoh, Kosuke, Univ. of Niigata, Brain Research Inst., Cntr Integrated Hum Brain Sci, Asahimachi-dori 1-757, Niigata, 951-8585, Japan

Ivanova, Valentina, ATCO Noise Management, HK, Engineering, 260 Holiday Inn Dr., Unit 1, Cambridge, ON N3C 4E8, Canada

Keaney, Diarmuid, ICAN Acoustics, St Marys Rd., Galway City, Ireland

Kinney, Norman, 1409 Rose St., Cape Girardeau, MO 63701-4621

Larose, Eric, LGIT, BP53, 38041 Grenoble Cedex 9, France

Le Bas, Pierre-Yves, 625 La bajada, Los Alamos, NM 87544

Leduc, Michel, 7765 Grand Rang, St-Hyacinthe, QC J2S 9B1, Canada

Lee, Seungkyu, Hankook Tire Co., R&D Ctr., NVH Research Team, 23-1, Jang-Dong, Yusung-Gu, Daejeon, 305-725, Korea

Legare, Luc, Montreal Public Transit, Fixed Equipment, 4692 Cyrille-Delage, Laval, QC H7K 3K5, Canada

Levert, Jason, 16974 Cedar Creek Ln, Noblesville, IN 46060

Levine, Jonathan, 11765 78th Ave., Allendale, MI 49401

Liberman, Mark, 3650 Spruce St., Philadelphia, PA 19104

Lieser, Eric, NAHB Research Ctr., Laboratory Services, 400 Prince Georges Blvd, Upper Marlboro, MD 20774

Lord, Sherman, Midlantic Technologies Group, 3191 Trewigtown Rd., Suite 100, Colmar, PA 18915
Luis, Joseph, ZEVEX, 4314 Zevex Park Ln., Salt Lake City, UT 84123
Malagoli, Armand, Airbus, Engineering Acoustics & Environment-EDEA(MO112/4), 316 Route de Bayone Cedex 09, Toulouse 31060, France
Manchester, Douglas, 6206 Goldeneye Court, Rocklin, CA 95765
Matthis, Kyle, 201 Surfrider Way, Apt. 2, Oceanside, CA 92054
McClure, Bob, BioSonics, Inc., Business Development, 4027 Leary Way NW, Seattle, WA 98107
McKellvie, Kent, Altermatt Associates Inc, 522 SW Fifth Ave. Suite 1200, Portland OR 97204
Milligan, Kelly, 5504 Old Spanish Trail, Bryan, TX 77807
Mintnert, Thomas, NTI Americas Inc., 7180 SW Fir Loop, Ste 210, Tigard, OR 97223
Miura, Masanobu, Ryukoku Univ., Media Informatics, 1-5, Yokotani, Oe-cho, Seta, Otsu, Shiga, 520-2194, Japan
Murray-Brown, Eugenie, 10 Mast Hill Rd., Hingham, MA 02043
Murray, Nathan, Univ. of Mississippi, NCPA, 1 Coliseum Dr., University, MS 38677
Niessen, Douglas, 212 Frankwood Dr., Folsom, CA 95630
Nobel, Marcus, 40712 SE Loudon Rd., Corbett, OR 97019
Otcenasek, Zdenek, Academy of Performing Arts, Music Faculty, Malostranski nam. 13, 118 00 Praha 1, Czech Republic
Park, Sang-Hoon, 3500 Greystone Dr., Apt. 212, Austin, TX 78731
Pedersen, Lars, Reson Inc., 100 Lopez Rd., Goleta, CA 93117
Pfitzinger, Hartmut, IPDS, CAU, Leibnizstr. 10, 24118 Kiel, Germany
Poo, Claudio, Universidad de Chile, IDIEM, Plaza Ercilla Poniente 883, 8370450 Santiago, Chile
Qin, Michael, NSMRL, Dive & Environmental Sim, NSMRL, Box 900, Groton, CT 06349-5900
Reinke, Robert, 1200 Rosemont Ave.. NW, Albuquerque, NM 87104
Rinehart, Hank, Radiance Technologies, Inc., 4 CR 2050, Oxford, MS 38655
Romeo, Victor, Ingenieria en Control Acustico, CEO, Rogelio Ugarte 1817, Santiago, 8360838 Santiago, Chile
Romeu, Jordi, Universidad Politecnica de Catalunya, Mechanical Engineering, Colom, 11 Etseiat, 08222 Terrassa, 08222 Barcelona, Spain
Ryan, William, Qipsis, LLC, 3815 Hobbs Rd., Greensboro, NC 27410
Saathoff, Shawn, 26526 Shoregrass Dr., Wesley Chapel, FL 33543
Saenz, Daniel, Pelton Marsh Kinsella, LLC, 5594 S Fort Apache, Suite 110, Las Vegas, NV 89148
Saltykov, Oleg, SIEMENS, R&D, 10 Constitution Ave., Piscataway, NJ 08854
Sato, Kenji, MItsubishi Precision Co., Ltd., Simulation Systems, 345 Kamimachiya, Kamakura City, 247-8505, Japan
Savarino, Thomas, 1087 Cherry Ave.., San Jose, CA 95125
Serafettinoglu, Hakan, BEKO Electronics AS, R&D Hardware Design, Beylikduzu Mevkii, Buyukcekmece, 34901 Istanbul, Turkey
Sergeev, Sergei, Michurinskiy Pr. 9/2-8, Moscow 119192 Russia
Shea, Kerry, 392 Monte Vista Way, Oceanside, CA 92057
Sheikh Nouri, Alireza, P.O. Box 14875-355, Shahre Ziba-Tehran-Iran, 14875 Tehran-Iran
Shiba, Hisashi, NEC Corporation, Radio Application Division, 1-10, Nisshion-cho, Fuchu, Tokyo, 183-8501, Japan
Stanton, Glenn, 6 Forest Ave., Old Tappan, NJ 07675
Stenson, Jim, OneSource Building Technology Inc., 961 Burke St., Winston-Salem, NC 27101
Stephany, Gary, Quest Technologies, 1060 Corporate Ctr. Dr., Oconomowoc, WI 53066
Storheim, Espen, Storheim, N-5938 Svegen, Norway
Su, Quang, State Univ. of New York, Mechanical Engineering, P.O. Box 6000, Binghamton, NY 13902-6000
Sweeney, James, Texas A & M Univ., 214 WERC, College Station, TX 77843
Talasch, Werner, Frankweg 18, Strasshof A-2231 Austria
Thorne, Peter, Proudman Oceanographic Laboratory, Joseph Proudman Building, 6 Brownlow St, Liverpool Merseyside L3 5SA, UK
Thunder, Stephen, Acoustic Associates, 1278 W. Northwest Hwy Ste 904, Palatine, IL 60067
Tomlinson, Seth, 28630 26th Ave. S., Federal Way, WA 98003
Unruh, Andrew, 15447 Stratford Dr., San Jose, CA 95124
Van Orden, Paul, City of Portland, 1900 SW 4th Ave.. Suite 5000, Portland, OR 97212
Visintini, Lucio, Via Monguelfo 4, 21100 Varese, Italy
Wei, Wei, 15848 NW West Union Rd., Apt.. #143, Portland, OR 97229
Whittington, Diana, US Fish and Wildlife Service, Division of Migratory Bird Man, 4401 N. Fairfax Dr., MS MBSP 4107, Arlington, VA 22203
Worland, Randy, Univ. of Puget Sound, Physics, 1500 N. Warner, Tacoma, WA 98416-1031
Wright, David, 17695 Bowdish Rd., Gregory, MI 48137
Yonovitz, Al, Univ. of Montana, Comm Sciences and Disorders, PMB 2012, 91 Campus Dr., Missoula, MT 59801
Zhang, Jingdong, Mobrey Ltd., Engineering, 158 Edinburgh Ave., Slough, SL1 4UE, UK
Zuleger, Garrett L., West Coast Environmental and Engineering, Dept. Engineering, 1838 Eastman Ave., Suite 200, Ventura, CA 93003

New Students

Abdul Hamid, Siti Zaleha, Faculty of Electrical Eng., Univ. Teknologi Malaysia, Utm,. Skudai, Johar, 81310 Malaysia
Agmon, Morag, Ben-Gurion Univ., Elect. and Comp. Engineering, PO Box 653, 84105 Beer-Sheva, Israel
Alizad, Azra, Mayo Clinic College of Medicine, Physiology and Biomedical Engineering, 200 1st St., SW, Rochester, MN 55905
Applegate, Chris, 527 W Beaver Ave., State College, PA 16801
Archer, Akibi, 1205 Maple Valley Ct., Union City, GA 30291
Bae, Youkyung, Univ. of Illinois at UC, Speech and Hearing Science, 901 South Sixth St., Champaign, IL 61820
Baek, Jong Dae, Stanford Univ., Mechanical Engineering, 102 Hoskins Ct., Apt.. 603, Stanford, CA 94305
Balaguru, RajAve..l, Stevens Inst. of Technology, Mechanical Engineering, Castle Point On Hudson, Hoboken, NJ 07030
Buck, Steven S., Univ. of Hartford, Box 3666, 200 Bloomfield Ave., West Hartford, CT 06117-1599
Cain, Jack D., 309 W. Kings Highway, Ephraim, NJ 08059
Chapman, Jacob, 5325 Turquoise Cir., Birmingham, AL 35242
Cheng, Eva, Univ. of Wollongong, SECTE (Bldg 35), Northfields Ave., Wollongong, 2522, Australia
Cheng, Roland, Univ. of Rochester, Electrical & Computer Eng, CSB526, RC Box 270231, Rochester, NY 14627
Cook, Brian, 840 Techwood Dr., Atlanta, GA 30313
De Vera, Jonathan, San Francisco State Univ., 2183 32nd Ave., San Francisco, CA 94116
Dennison, Heeyeon, 1711 East-West Rd., #825, Honolulu, HI 96848
Desherlia, Andrew, Columbia College Chicago, Audio Arts and Acoustics, 6643 N Newgard, West Chicago, IL 60626
Dicanio, Christian, PMB 332, 2425B Channing Way, Berkeley, CA 94704
Dillon, Jeremy, 22 Vancouver St., St. John's, NF A1A 2R6, Canada
Drott, Emily, 14535 Carolcrest Dr, Houston, TX 77079
Dun, Yujie, Xi'an Jiaotong Univ., Information & Communication Eng., Xi'an Jiaotong Univ., 710049 Xi'An, P. R.China
Earnhart, Nicholas, 1002 Curran St. NW, Atlanta, GA 30318
Ely, Timothy, 116 Plum Brook Rd., Flemington, NJ 08822
Feffer, Yvonne, 1500 Lawrence Ln., Northbrook, IL 60062
Foertsch, Kevin, 1972 Grand View Dr., Jasper, IN 47546
Ganseman, Joachim, Maantjessteenweg 160, B2170 Merksem, Belgium
Gereoudakis, Nikolaos, 69 Cropley St., London, N1 7JB, UK
Giannakopoulou, Anastasia, Brunel Univ., Psychology Dept., 157 Marie Jahoda, Uxbridge, UB8 3PH, UK
Gillett, Lyric, 1821 Harvard St., Houston, TX 77008
Gotzke, Carrie, Univ. of Alberta, Speech Pathology and Audiology, 2-70 Corbett Hall, Edmonton, AB T6G 2G4, Canada
Grigoletti, Steven, 3547 N. Marshfield, Chicago, IL 60657
Hawco, Colin, 387 ERB St., West, Apt. 11, Waterloo, ON H4G 2W6 Canada
Hazen, Elliott, 135 Duke Marine Lab Rd., Beaufort, NC 28516
Ho, Cheng-Yu, National Yang-Ming Univ., Inst of Biomedical Engineering, No 155 Li-Nong ST, Sec 2, Taipei 112, Taiwan
Holmes, Blythe, 1406 W. Harrison St. #3C, Chicago, IL 60607
Hong, Kwanwoo, 140 Airport Rd., Apt. 6, West Lafayette, IN 47906

Hou, Kang, Purdue Univ., Mechanical Engineering, Ray W. Herrick Lab, Purdue Univ, 140
S. Martin Jischke Dr., West Lafayette, IN 47906
Hristev, Ivailo, jk.Liulin 2 bl.238 vh.B, Sofia, Bulgaria
Hsu, Timothy, Georgia Tech, 251 10th St., NW, Bldg. A506, Atlanta, GA 30318
Jeon, Hae-Sung, Univ. of Cambridge, Dept. of Linguistics, 672 King's College, Cambridge, CB2 1ST, UK
Kankey, Andrew, 10 Vairo Blvd., Apt. 25C, State College, PA 16803
Kehr, Thomas, 717 Central Park Blvd., Port Orange, FL 32127
Kim, Taekyoon, KAIST, Mechanical Engineering, Rm. 5121, ME Building, Science Town, Daejeon, 305-701, Korea
Kim, Dongjoon, 233 E. 32nd St., Apt..# 3C, New York, NY 10016
Klymko, Victor, P.O. Box 8646, University, MS 38677
Konnai, Ramya, Communication Disorders, 200 Health Ctr. Building, Bowling Green State Univ., Bowling Green, OH 43402
Krueger, David, 4888 Mountain View Cir., Highland, UT 84003
Kucera, Sam, The Univ. of Tulsa, Physics and Engineering Physics, 800 South Tucker Dr., Tulsa, OK 74104
Lamkanfi, Ebrahim, Sint-Pietersnieuwstraat 41, Mechanical Construction and Pr, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium
Lansford, Kaitlin, 2514 W. Carter Dr., Tempe, AZ 85282
Laurin, Scott, Penn State, Acoustics, 747 A Whitehall Rd., State College, PA 16801
Legendre, Susan, 7762 S. Trumpet Vine Ave.., Tucson, AZ 85747
Lehmann, Kenna, 2111 Quail Ranch Ct., Elverta, CA 95626
Levada, Thiago, Rua Padre Carvalho, n 200, Sco Paulo, 05427020, Brazil
Luckey, Robert, 920 Greenland Dr., Apt.. 312, Murfreesboro, TN 37130
Luneau, Jean-Marc, Aalborg Univ., ES-MISP, Niels Jernes Vej 12, A6-302, 9220 Aalborg, Denmark
Marco, Todd, Penn State Univ., Electrical Engineering, 135 Glenville Ave., Allston, MA 02134
Marin, Jorge, 301 10th St., NW, 517C, Atlanta, GA 30318
McLaughlin, Bonnie, 140 S 800 E #6, Salt Lake City, UT 84102
Mehmood, Asif, P.O. Box 4206, University, MS 38677
Millar, Kendall, 8090 Franklin Rd., Hagerstown, IN 47346
Nandlall, Sacha, BUBL Oxford, Engineering Science, 17 Parks Rd., Oxford, OX1 3PJ, UK
Nelson, Jonah C., 582 Hanson Rd., Ellensburg, WA 98926
Neves, Silvana, Gatty Marine Lab., Univ. of St Andrews, St Andrews, KY16 8 LB, UK
Nevitt, Rachelle, Brigham Young Univ., Communication Disorders, 136 TLRB, Provo, UT 84602
Pate, John, 82 East Maynard Ave., Columbus, OH 43202
Pizarek, Thomas F., Wabash College, Physics, P.O. Box 352, Crawfordsville, IN 47933
Poling, Jacob, 1801 W. North Ave., Apt. 3F, Chicago, IL 60622
Reichling, Michelle, CTDG, Acoustics Consulting, 901 N. Calvert St., 2nd Floor Apt. B, Baltimore, MD 21202
Roe, Matthew, 267 Meigs St., Apt. 1, Rochester, NY 14607
Rosenbaum, Joyce, 315 S. Allen St,. Apt. 523, State College, PA 16801
Ross, Tristie, 1 Miramar St. #929557, La Jolla, CA 92092
Rozinek, Sarah, 12033 Madrona, San Antonio, TX 78245
Saarni, Tuomo, Univ. of Turku, Information Technology, Joukahaisenkatu 3-5, 20014 Turku, Finland
Sato, Yuji, Hanabatake 3-11-4 402, Tsukuba, 3003261, Japan
Schoeman, John, 37 Promontory Ridge Dr., Las Vegas, NV 89135
Schoenfelder, Vinzenz, TU Berlin, Sekr. FR 6-4, Franklinstrasse 28/29, 10587 Berlin, Germany
Schultz, Jim, 10317 W. Cascade Dr., Franklin, WI 53132
Shaver, Mark, Wichita State Univ., CSD, 1845 Fairmount, Box 75, Wichita, KS 67260-0075
Shaw, Jason, 690 10th St., Brooklyn, NY 11215
Shen, Jing, 7510 Charmant Dr., #726, San Diego, CA 92122
Shorter, Mellonise, 2133 W. 111th St., Los Angeles, CA 90047
Sidaras, Sabrina, 4301 Sloan Square, Atlanta, GA 30329
Simko, Peter, IL Inst. of Technology, ECE, 3301 S. Dearborn, Chicago, IL 60616
Sims, Ricky, 4216 Leesburg Rd, Lt 3, Hopkins, SC 29061
Sirpotdar, Sanket, 5200 Anthony Wayne Dr., Apt. 515, Detroit, MI 48202
Sjerps, Matthias, Max Planck Inst. for Psych., Comprehension, Wundtlaan 1, 6525 XD Nijmgen, The Netherlands
Skoug, Ryan, 6903 Oak Plaza, #603, Omaha, NE 68106
Small, Christopher S., Carthage College, Physics, 2001 Alford Park Dr., Kenosha, WI 53140
Stepp, Cara, MGH Voice Ctr., 1 Bowdoin Sq—11th Flr., Boston, MA 02114
Syed, Moinuddin, C/O Lena Magnusson, Blekinge Tekniska Hogskolan, Campus Grasvik, 37179 Karlskrona, Sweden
Taylor, Rebecca, 36 The Old Village, Huntington, York, YO32 9RB, UK
Toland, Andrew, 12230 SW Foothill Dr., Portland, OR 97225
Treichel, Toni, Neuenburger Weg 19, 04357 Leipzig, Germany
Turner, Katrine, Fynsgade 11, st th, 8000 Aarhus C, Denmark
Tyson, Na'Im, 2897 Neil Ave., Apt. 380A, Columbus, OH 43202
Vernaleo, Beth, 4333 Rowalt Dr., Apt. #302, College Park, MD 20740
Wake, Melissa, 18912 Lister Ln., Huntington Beach, CA 92646
Winter, Adam Z., 32333 227th, SE, Kent, WA 98042
Wirthlin, Morgan, 5417 S. Ellis Ave., #2, Chicago, IL 60615
Yano, Benjamin, Columbia College Chicago, Acoustics, 2 East 8th St., Chicago, IL 60605
Zepidou, Georgia, Flat 97, North Block, 5 Chicheley St., London, UK
Zucker, Michael, 70 White Oak Ln., Etters, PA 17319

New Electronic Associates

Bockhorst, Rhea, Bockhorst Consulting, 111 Riverside Ln., Marietta, OH 45750
Bourdon, Raphael, Residence de Lann Kerguipp, 29350 Moelan Sur Mer, France
Brantley, Merry, Naval Branch Health Clinic, Audiology, One Ayres Cir., Bldg. H-1, Portsmouth Naval Shipyard, Portsmouth, NH 03804-5000
Chan, Dominic, ASTRI, ECE, Bio-Informatics Centre, 3/F, 2 Science Park West Ave., Hksp, Shatin HKSAR, Hong Kong
Chen, Xiaofei, Univ of Sci and Tech of China, Sch of Earth and Space Sci, 96 Jin-Zhai Rd., 230026 Hefei, Anhui Prov., P.R. China
Choiniere, Marc, 623 Admiral Dr., Unit 303, Annapolis, MD 21401
Clark, Stewart, 508 Redlands, Newport Beach, CA 92663
Cornelisse, Leonard, Unitron Hearing, R&D, 20 Beasley Dr., P.O. Box 9017, Kitchener, ON N2J 4Y9, Canada
Enomoto, Seigo, NICT/ATR-SLC, ASR, 2-2-2 Hikari-dai, Seika-cho, Souraku-gun, Kyoto-Fu, Japan
Krajcarz, Fabien, Gamba Acoustique, industrie et environnement, buro parc 2, rue de la decouverte, 31676 Labege, France
Kuo, Yu-Ching, Taipei Municipal U. of Education, Special Education, No. 1, Ai-Guo West Rd., Taipei 10042, Taiwan
Lafrance, Patrick, Environnement Illimiti, 1453 Saint-Timothie, Montreal, QC H2 L 3N7, Canada
Leclaire, Philippe, LRMA-ISAT, 49 rue Mademoiselle Bourgeois, B.P.31, 58027 Nevers, France
Levelt, Clara, Leiden Univ., Linguistics, PO Box 9515, 2300 RA Leiden, The Netherlands
Mcneur, Robyn, Phonak NZ Ltd, PO Box 33-349, Takapuna, Auckland, New Zealand
Mills, Alan, 188 Calle Madrid, St. Augustine, FL 32086
Nielsen, Leif, ISOITC43 & IECIT29, Kollegievet 6, Charlottenlund DK Denmark
Norisuye, Tomohisa, Kyoto Inst. of Technology, Dept. of Macromolecular S, Matsugasaki, Sakyo-ku, Kyoto, Kyoto, 606-8585, Japan
Ravitch, Vladimir, Nexense, R&D, Hakishon 1, 81220 Yavne, Israel
Reilly, Sean, Navy Warfare Development Cmd, Modeling and Simulation, 686 Cushing Rd., Newport, RI 02881
Rick, David, Hach Company, Flow and Sampling BU, 5600 Lindbergh Dr., Loveland, CO 80538
Sawada, Kouichi, NRIFE, FRA, 7620-7 Hasaki, Kamisu, Ibaraki, 314-0408, Japan
Secail, Mathieu, PARROT, Signal Processing, 174 quai de Jemmapes, 75010 Paris, France
Stewart, John, 1201 S. Ohio St., Martinsville, IN 46151
Suzuki, Kenji, Konicaminolta Medical & Graphics, R&D Ctr. HC No.5, 1 Sakurac-machi, Hino-shi, Tokyo, 191-8511, Japan
Swarts, J. Douglas, Children's Hospital of Pittsburgh, Pediatric Otolaryngology, 3705 Fifth Ave., Pittsburgh, PA 15213

Tarabini, Marco, Politecnico di Milano, Mechanics, Via m d'oggiono 18/a, Lecco, Italy

Wu, Yu-Hsiang, Univ. of Iowa, Speech Pathology & Audiology, 119 SHC, Iowa City, IA 52242

Electronic Associates

Lee, Ki-Seung, Konkuk Univ., Electronic Engineering, Kwangjin-Gu Hwayang-dong, Seoul, 143-701, Korea

Lim, Jun-Seok, Sejong Univ., Dept. of Electronics Eng., 98 Kwangjin Kunja, Seoul, 143-747, Korea

Rodrigues, Cleide, IPqM, Rua Ipiru, 2, Rio De Janeiro, 21931-095, Brazil

Saikia, Siddhartha, New 5/Old 18, Vasudevan Nagar, Thiruvanmiyur, Chennai 600041, India

Tanzil, Leo, PT Kairos Multi Jaya, Design and Engineering, Agung Perkasa Blok K1 no 4, Jakarta 14350, Indonesia

Yiow, Chong Lyat, dB Kawal Sdn Bhd, 9-2 Jalan USJ 1/1A, Subang Jaya, 47620 Selangor, Malaysia

Associates Elected Members

C. J. Baillie, G. M. Barker, T. C. Bent, D. T. Bradley, J. T. Cheng, D. Delaney, A. C. Faszer, C. P. Giegold, D. Jurevicius, K. Lefkowitz, J. J. Lopez, L. Maffei, B. G. Marston, K. Nishi, M. K. Norris, C. H. Oppenheimer, R. C. Schrag, D. R. Schuette, J. C. Shearer, R. Smiljanic, J. M. Stout, E. P. Stride, S. J. Stulgin, Z. Su, J. E. Summers, C-T. Tan, B. G. A. Van Zanten, A. M. Yazdannivaz, L. Zoontjens

Electronic Associates Elected Members

J. F. Abbott, W. Chu, G. S. Sparks

Student to Associate

S. Lee, R. S. Scott

Associate to Electronic Associate

A. N. Miller, B. K. Mukherjee

Student to Electronic Associate

K. Jambrosic

Resigned

S. Saito, *Fellow*

G. A. Mangiante, H. Ermert, S. Wisotsky, Sr., A. N. Woodger, *Members*

J. Birli, T. Busch, A. Ebrahim, D. Frederick, D. A. Hall, *Associates*

B. W. Lewis, P. Nicholson, M. Ziolkowski, *Electronic Associates*

G. P. Allen, C. Mathews, A. Messina, K. Reindl, *Students*

Deceased

H. E. Bass, D. M. Chase, R. R. Goodman, J. M. Harrison, H. D. McNiven, E. Mikeska, J. Pope, *Fellows*

O. O. Gruenz, O. Jacobsen, F. R. Mallalieu, W. A. Manson, P. W. Nye, M. W. Skinner, G. R. Southworth, J. Stepanek, M. S. Weinstein, *Members*

H. K. Cheng, *Associate*

Dropped

John Bouyoucos, Homer Bucker, W. Jack Hughes, *Fellows*

Donald Allen, Yoav Arieh, Rune Axelsson, Chetlur Balachandran, Charles Barker, Arjan Bosman, Geng Chen, Kuo-Tsai Chen, Cheri Deng, Samantha Dugelay, Kenneth Erikson, Alice Faber, Harold Farr, Ching Feng, Levassort Franck, Robert Gilkey, Jingfeng Guan, Zaegyoo Hah, Satoshi Imai, Soon Jarng, Paul Joppa, Seong-Woo Kang, John Kesner, Damir Khismatullin, John Kingston, Vikram Kinra, Masanobu Kumada, Jonathan Lally, Paul P Lee, Jean-Paul Lefevre, Xun Lei, Yury Lysanov, Neil MacMillan, Lawrence Maga, Margaret Mortz, Anders Nilsson, Ronald Nishi, Stephen O'Leary, Kofi Owusu-Anash, Anand Parthasarathi, Scott Perron, Oguz Poroy, Robert Porter, Zhang Qiong, Jacob Ross, C Guy Rudisill, Michael Salameh, August Schick, Mitchell Shipley, Dick Simons, Kirk P Sullivan, Jan Thoen, Sergio Vera, Yoshio Yamasaki, Duen Hsi Yen, Jane Young, *Members*

Marcus Adrian, George Antonelli, Zeev Ariel, Raymond Bargain, David Basinger, Paul Beard, Michael Beauvois, Steven Benzel, Patricia Berge, Michael Bird, Lance Bischoff, Gabriel Bitton, Yolande Bonin, Edward Booth, Frank Boyle, Marc Bracken, Richard Brander, Merry Brantley, Michel Brissaud, Stephen Burcsak, Walter Carney, Samuel Carroll, Maureen Casper, Silvio Cavuto, Henrietta Cedergren, Jan Charles-Luce, Monita Chatterjee, B Chaudhary, Michael Chial, Bok Kyoung Choi, Chul-Hee Choi, James Christensen, Joseph Cichock, Marsha Clarkson, Jeffry Coady, Marc Colen, Matthew Conti, John Crawford, Charles Dailey, Praven Dala-krishna, Steven Davis, Maarten De Hoop, Francine Desharnais, Anthony Destephen, Joseph Digerness, Christine Dixon-Ernst, Michael Donaldson, Marshall Downing, Michael Dungan, Antonio Dura, Glenn Eckenrode, Sharry Edwards, Nathan Eide, Jeff Eldredge, Fernando Elizondo, Wesley Elsberry, Gulam Emadi, Hari Eswaran, Marta Etcheverry de Milou, Susan Fallon, David Feinberg, Sidney Fels, Marcia Ferguson, Guild Fetridge, David Fisher, Matthew Fitzgerald, Jane D Flood, Thomas Folegot, Kenneth Frampton, Jon Francine, Eric Frenz, Kenneth Frere, Elaina Frieda, Rod Gervais, Sean Gilmour, William Graham, Thomas Gross, Hongkai Guo, Gina Hagler, Michael Hall, Ibrahim Hallaj, Wali Hamidzada, Ruijing Han, Brett Hanes, Anming He, Kyle Hollman, Ethan Honda, Derek Houston, Lianjie Huang, Michael Huelskemper, Farkhanda Jabeen, Paul Johnson, Hirofumi Kamata, George Kapodistrias, Eric Karas, David Kastak, Alan Kefauver, Eui Jun Kim, Hee-Sun Kim, Jin Kim, Ok Kim, Ryan Knox, Tetsunori Kobayashi, Narayanan Kuthirummal, Emily Lally, Chantal Laroche, Andrew LaRow, James Larue, Scott Lauback, Johan Leander, Stanley Lemon, Weili Lin, Kenneth Line, Virginia Lipscy, Arthur Lobo, Jeffry Long, Li Ma, James Mahshie, Chi-Min Mao, Kyle Martini, Russell Mason, George McCall, Ryan McCormick, Duncan McGehee, Bin Mi, Julian Morgan, David Moulton, Murray Munro, Misty Nelson, Kok-Hwee Ng, Kevin O'Neill, John Ochs, Ossi Ojanen, Pedro Olalla, Boyd Osgood, Daejin Park, Teralandar Parthasarathy, Richard Paschke, Aaron Peters, William Pielemeier, Mario Pinto, Michael Pitermann, Anthony Puckett, Jinyu Qian, Shengping Qin, Edmond Rambod, Paul Raven, Daniel Russell, Bruce Sabol, Mal Sacks, Carl Sandrock, Joseph Santos-Sacchi, John Schneider, David Senzig, Kim Seung, Karen Shandera, Kiran Shanker, Michael Shatalov, Anthony Sinclair, Balbir Singh, Barry Slansky, Dahlia Sokolov, Rotimi Soremekun, Renato Spagnolo, Abigail Stefaniw, Lei Sui, Sanjay Sundermurthy, Marija Tabain, Sumiko Takayanagi, Haruo Tanaka, Erik Teixeira, Maria Terra, Mark Trainer, Danielle Tryanski, Joel Tuss, Emmanuel Tzekakis, Hisashi Uematsu, Apparao Vaddempudi, James Valentine, Harry Van Trees, Duane Voskuil, Magnus Wahlberg, Timothy Walsh, Joseph Warda, Kevin Warner, Joseph Whalen, James White, T Wilson, Jianqiang Xiao, Alfred Yalaju, Xinmai Yang, Michael Yargus, Ueda Yasutaka, Dae Hee Youn, Alan Yu, Erika Zettner, Xian Zhang, Jalal Zia, *Associates*

Shani Abada, Ali Akar, Lilian Aroutiounian, Junghwa Bahang, Helen Bailey, Aditya Balani, Neil Bardhan, Alan Beaudry, Daniel Beckmann, David Benson, Sarah Benson-Amram, Aurelie Bidet-Caulet, Paul Black, Sebastian Boeser, Ann Bogacka, Derek Bonner, Rose Bradley, Kaca Bradonikic, John Brady, Judson Brown, Julian Burgos, Cecilia Camacho-Ramirez, Dario Carmignani, Julie Case, Amanda Cerka, Lan Chen, Li-Jen Chen, Rui Cheng, Danielle Cholewiak, Hyunjoo Chung, Yoojin Chung, Shaun Clair, Linde Clark, Matthew Colvard, Justin Crosby, James Dalberg, Erdenebat Dashtseren, Celeste DeFreitas, Alana Deloach, William DeMeo, Nicholas DiZinno, Andrew Draudt, Alexandra Dunn, Mark Ehrlich, Mona El-Kady, Vahit Erdener, Joe Erickson, Jason Ernst, Zhaleh Feizollahi, Kazuhiro Fujisawa, Melissa Fuller, Elizabeth Furr, Vincent Galea, Anthony Gerard, James German, Stefan Goetze, Amanda Graves, Justin Gregg, Melissa Gregg, Rabindranath Gubbala, Mustafa Gunduz, Christopher Hamilton, Jamie Hamilton, Ryan Hanke, Nicholas HAve..ns, Christopher Hennes, Alexis Hervais-Adelman, Harshall Hilton, Connie Ho, Jason Holmes, Alex Hornecker, Miguel Horta, I-Hui Hsieh, Harvey Hudson, Eli Hughes, Jonathan IIoreta, Ananth Iyer, Sandip Jassar, Justin Jaworski, Haisheng Jiang, Ackland Jones, Guenevere Jones, Aditya Joshi, Wonmo Kang, Elizabeth Kennedy, Ryan Kenny, Sungyoung Kim, Yunjung Kim, Andrea Kosko, Sandra Kreidler, Cenny Kuesal, Danielle Kumpulanian, Sangyeop Kwak, Cheryl Lacsamana, Mario Landera, Florian Lanz, Christophe Laplanche, Connie Lee, Hotaik Lee, Jaehyung Lee, Ok Kyun Lee,

Erica Lenhart, Anne-Marie Lerner, Lucas Lezamiz, Naihsin Li, Andreas Link, Sheng Liu, Bryce Lobdell, Elbereth Lorenz, Katharine Love, Bert Maat, Mark MacLeod, Shawn Makepeace, Caroline Maleke, Coleman Martin, Karen Maruska, Candice McCarty, Anthony McCreery, Karla Melendez, Hongyu Miao, Jacob Migliazzo, Lucas Monsalve, Terri Morrison, Marie Muller, Nathan Naluai, Kuniko Nielsen, David Nutter, Jaclyn Ocumpaugh, Anne Olmstead, Douglas Orellana, Nathan Pape, Akshay Patel, Allison Paumen, Isaac Penny, Jonathan Pincas, Eleni Pinnow, Tracy Pogal-Sussman, S Prakash, Ying Que, Alexander Queenan, Jorge Quijano, Sreedivya Radhakrishnan, Erik Rodriguez, Deanna Rogers, Sarah Rohrkasse, Julian Romero, Jessica Rossi-Katz, Elin Roverud, Yohei Sakata, Flore Samaran, Ashwinkumar Sampathkumar, Hamid Satar-Boroujeni, Harsha Sathyendra, S Satyavarta, Jennifer Saulus, Jacob Scarpaci, William Schenck, Rodrigo Segnini, Jonathan Sellon, Hojune Seo, Paul Shepherd, Yoon Shin, Douglas Sladen, Jeremy Slagley, Charles Smith, Marc Smith, Moira Smith, Noelle Smith, Connie So, Valerie Speciale, Mychal Spencer, Marko Stamenovic, Michael Stone, Woo Sung, Ryuji Suzuki, Alexandra Symmonds, Navid Taghvaei, Jonathan Thierman, Barry Threw, Roxana Tibrea, Henrique Tono, Justin Trout, Manolis Tsakiris, Juan Tu, Emily Van Ark, Hillary Vargas, Kristin Vasil, Manish Velankar, Navin Viswanathan, Oveal Walker, Weixiong Wang, Kevin Webster, Joanna Webster Tampas, Andrew Whelan, Jarrod Whittington, Colby Wilkason, Scott Woolston, Tianming Wu, Tina Yack, Jie Yang, John Zeman, Shukui Zhao, *Students*

Barbara Acker-Mills, Matt Allen, Richard Anderson, Kazuhiko Anraku, Scott Backhaus, Christopher Bader, Diane Blackwood, William Blank, Michael Blommer, Cormac Booth, Meir Botbol, Giola Bottai, Raymond Brueckner, Jonas Brunskog, Bruce Bullock, Rose Burkholder, Christopher Busch, John Busenitz, Densil Cabrera, Brian Capleton, Raymond Carroll, Cheong Cheolung, Ernest Cholakis, Claude Cohen-Bacrie, Rupert Cox, Penny Davis, Konstantinos Douligeris, Roberto Elizalde-Dounce, Stephen Gauld, Noel Gavenvelle, David Gerhard, Bernd Gottinger, Massimo Grassi, Doug Greenlee, Amitava Gupta, Steven Haddock, Mark Hayner, Ricardo Higuti, Caroline Huang, Gwoanchul Hyun, Kiyoshi Ishikawa, Armando Jaramillo Legorreta, Ossie Jones, Sung Kim, Tobias Kirchner, Sabrina Kleinknecht, Ying-Yee Kong, Courtney Lane, Albert-Georg Lang, Scott Lee, Mark Lehman, Charles Lipscomb, David Lopez, Ramon Lopez-Vidal, Sanjay Mahajan, James Mansky, Xavier Maruyama, Michael Mease, James Meyers, Ortega Miguel, Carmelo Militello, Maxwell Miller, Jan MyoungWoo, Mandayam Naresh, Areti Okalidou, Brian Oppegaard, Amita Pal, Chris Pannell, Jangyeon Park, Rupal Patel, Matthew Pettersson, Steve Puffenberger, Robert Qiu, Robert Quinn, Sean Ramprashad, Davilla Riddle, Steven Rines, Laia Romero, Samuil Rybak, Marc Sabat, Ariel Salomon, Jihene Serkhane, Sergey Shabad, Gaurav Shah, Carl Simpson, Joel Snyder, Janet Stack, Andrea Szigetvari, Philip Thomson, Richard White, Kainan Wong, Shi-Chang Wooh, David Wright, Gangyi Yang, Wen-Bin Yang, Miguel Zenker, *Electronic Associates*

Fellows	909
Members	2136
Associates	2756
Electronic Associates	682
Students	1088
	7581

ACOUSTICAL STANDARDS NEWS

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

Paul D. Schomer, Standards Director
Schomer and Associates, 2117 Robert Drive, Champaign, IL 61821 [Tel.: (217) 359-6602; Fax: (217) 359-3303; e-mail: schomer@schomerandassociates.com]

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

Standards Meetings Calendar—National

The 156th ASA meeting will take place in Miami, Florida, at the Doral Golf Resort and Spa, 10–14 November 2008. More information, including online registration, hotel information and a Call for Papers, can be found on the ASA website at http://asa.aip.org/meetings.html.

• ***Monday, November 10, 2008***

– 7:00 p.m.–10:00 p.m.—ASACOS Steering Committee

• ***Tuesday, November 11, 2008***

– 7:30 a.m.–9:00 a.m.—ASA Committee on Standards (ASACOS) Meeting of the Committee that directs the Standards Program of the Acoustical Society

– 9:15 a.m.–10:30 a.m.—Standards Plenary Group (includes U.S. TAGs to ISO/TC 43, TC 43/SC 1, ISO/TC 108 and its subcommittees, and IEC/TC 29)

– 11:00 a.m.–12:45 p.m.—Accredited Standards Committee S12, Noise

– 2:15 p.m.–3:30 p.m.—Accredited Standards Committee S1, Acoustics

– 3:45 p.m.–5:00 p.m.—Accredited Standards Committee, S2, Mechanical Vibration and Shock

• ***Wednesday, November 12, 2008***

– 9:00 a.m.–10:30 a.m.—Accredited Standards Committee S3, Bioacoustics

– 10:45 a.m.–12:00 noon—Accredited Standards Committee S3/SC1, Animal Bioacoustics

Standards Meetings Calendar—International

• ***3–7 November 2008***

ISO/TC108 Mechanical vibration, shock and condition monitoring, is meeting in St. Louis, Missouri, USA, along with subcommittee **ISO/TC108/SC 3** Use of calibration of vibration and shock measuring instruments.

• ***30 March–3 April, 2009***

ISO/TC108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures will meet in London, UK.

Approval of Reaccreditation!

ASC S1—Acoustics; ASC S2—Mechanical Vibration and Shock; ASC S3—Bioacoustics; ASC S12—Noise

ASA Standards is happy to report that ANSI's Executive Standards Council has approved the reaccreditations of the above Accredited Standards Committees (ASCs) under revised operating procedures for documenting consensus on proposed American National Standards, effective August 13, 2008.

Take Note!

In July 2008 Accredited Standards Committee S2, Mechanical Vibration and Shock, approved the following erratum to ANSI S2.8-2007 *American National Standard Technical Information Used for Resilient Mounting Applications*:

Erratum to ANSI S2.8-2007 American National Standard Technical Information Used for Resilient Mounting Applications

ANSI S2.8-2007 contains the following typographical error. It occurs on page 8 in Clause 5.2, Equation (6)

The equation, as published, reads:

$$2\pi f_n = \sqrt{\frac{k}{g}} = \sqrt{\frac{kg}{W}} = \sqrt{\frac{g}{X_{st}}} \tag{6}$$

The equation given is incorrect and should be:

$$2\pi f_n = \sqrt{\frac{k}{m}} = \sqrt{\frac{kg}{W}} = \sqrt{\frac{g}{X_{st}}} \tag{6'}$$

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

A listing of the Accredited Standards Committees, their Working Groups, their Chairs (listed in parenthesis), and published standards; Chair and Vice Chair of the ASA Committees on Standards (ASACOS); and the U.S. Technical Advisory Group (TAG) Chairs for the International Standards Committees are given here for reference:

 © 2008 Acoustical Society of America

ACCREDITED STANDARDS COMMITTEE ON ACOUSTICS, S1

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

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

S1 Working Groups

S1/Advisory—Advisory Planning Committee to S1 (P. Battenberg, Chair; R.J. Peppin, Vice Chair);
S1/WG1—Standard Microphones and their Calibration (V. Nedzelnitsky);
S1/WG4—Measurement of Sound Pressure Levels in Air (E. Dunens);
S1/WG5—Band Filter Sets (A.H. Marsh);
S1/WG9—Calibration of Underwater Electroacoustic Transducers (R.M. Drake);
S1/WG17—Sound Level Meters and Integrating Sound Level Meters (G.R. Stephany);
S1/WG19—Insertion Loss of Windscreens (A.J. Campanella);
S1/WG20—Ground Impedance (Measurement of Ground Impedance and Attenuation of Sound Due to the Ground) (K. Attenborough, Chair; J. Sabatier, Vice Chair);
S1/WG22—Bubble Detection and Cavitation Monitoring (Vacant);
S1/WG26—High Frequency Calibration of the Pressure Sensitivity of Microphones (A. Zuckerwar);
S1/WG27—Acoustical Terminology (J.S. Vipperman).

S1 Inactive Working Groups

S1/WG15—Noise Canceling Microphones (R.L. McKinley);
S1/WG16—FFT Acoustical Analyzers (R.J. Peppin);
S1/WG21—Electromagnetic Susceptibility (EMS) of Acoustical Instruments (J.P. Seiler);
S1/WG24—Design Response of Weighting Networks for Acoustical Measurements (G.S.K. Wong);
S1/WG25—Specification for Acoustical Calibrators (P. Battenberg).

S1 Standards on Acoustics

ANSI S1.1-1994 (R 2004) American National Standard Acoustical Terminology
ANSI S1.4-1983 (R 2006) American National Standard Specification for Sound Level Meters (This Standard includes **ANSI S1.4A-1985 (R 2001)** Amendment to ANSI S1.4-1983.)
ANSI S1.6-1984 (R 2006) American National Standard Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements
ANSI S1.8-1989 (R 2006) American National Standard Reference Quantities for Acoustical Levels
ANSI S1.9-1996 (R 2006) American National Standard Instruments for the Measurement of Sound Intensity
ANSI S1.11-2004 American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters
ANSI S1.13-2005 American National Standard Measurement of Sound Pressure Levels in Air
ANSI/ASA S1.14-1998 (R 2008) American National Standard Recommendations for Specifying and Testing the Susceptibility of Acoustical Instruments to Radiated Radio-Frequency Electromagnetic Fields, 25 MHz to 1 GHz
ANSI S1.15/Part 1-1997 (R 2006) American National Standard Measurement Microphones, Part 1: Specifications for Laboratory Standard Microphones
ANSI S1.15/Part 2-2005 American National Standard Measurement Microphones, Part 2: Primary Method for Pressure Calibration of Laboratory Standard Microphones by the Reciprocity Technique
ANSI S1.16-2000 (R 2005) American National Standard Method for Measuring the Performance of Noise Discriminating and Noise Canceling Microphones
ANSI S1.17/Part 1-2004 American National Standard Microphone Windscreens—Part 1: Measurements and Specification of Insertion Loss in Still or Slightly Moving Air
ANSI S1.18-1999 (R 2004) American National Standard Template Method for Ground Impedance
ANSI S1.20-1988 (R 2003) American National Standard Procedures for Calibration of Underwater Electroacoustic Transducers
ANSI S1.22-1992 (R 2007) American National Standard Scales and Sizes for Frequency Characteristics and Polar Diagrams in Acoustics
ANSI S1.24 TR-2002 (R 2007) ANSI Technical Report Bubble Detection and Cavitation Monitoring
ANSI S1.25-1991 (R 2007) American National Standard Specification for Personal Noise Dosimeters
ANSI S1.26-1995 (R 2004) American National Standard Method for Calculation of the Absorption of Sound by the Atmosphere
ANSI S1.40-2006 American National Standard Specifications and Verification Procedures for Sound Calibrators
ANSI S1.42-2001 (R 2006) American National Standard Design Response of Weighting Networks for Acoustical Measurements
ANSI S1.43-1997 (R 2007) American National Standard Specifications for Integrating-Averaging Sound Level Meters

ACCREDITED STANDARDS COMMITTEE ON MECHANICAL VIBRATION AND SHOCK, S2

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

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

S2 Working Groups

S2/WG1—S2 Advisory Planning Committee (A.T. Herfat, Chair; R.L. Eshleman, Vice Chair);
S2/WG2—Terminology and Nomenclature in the Field of Mechanical Vibration and Shock and Condition Monitoring and Diagnostics of Machines (D.J. Evans);
S2/WG3—Signal Processing Methods (T.S. Edwards);
S2/WG4—Characterization of the Dynamic Mechanical Properties of Viscoelastic Polymers (W.M. Madigosky, Chair; J. Niemiec, Vice Chair);
S2/WG5—Use and Calibration of Vibration and Shock Measuring Instruments (D.J. Evans, Chair; B.E. Douglas, Vice Chair);
S2/WG6—Vibration and Shock Actuators (G. Booth);
S2/WG7—Acquisition of Mechanical Vibration and Shock Measurement Data (B.E. Douglas);
S2/WG8—Analysis Methods of Structural Dynamics (M. Mezache);
S2/WG9—Training and Accreditation (R. Eshleman, Chair; D. Corelli, Vice Chair);
S2/WG10—Measurement and Evaluation of Machinery for Acceptance and Condition (R.L. Eshleman, Chair; H. Pusey, Vice Chair);
S2/WG10/Panel 1—Balancing (R.L. Eshleman);
S2/WG10/Panel 2—Operational Monitoring and Condition Evaluation (R. Bankert);
S2/WG10/Panel 3—Machinery Testing (R.L. Eshleman);
S2/WG10/Panel 4—Prognosis (A.J. Hess);
S2/WG10/Panel 5—Data Processing, Communication, and Presentation (K. Bever);
S2/WG11—Measurement and Evaluation of Mechanical Vibration of Vehicles (A.F. Kilcullen);
S2/WG12—Measurement and Evaluation of Structures and Structural Systems for Assessment and Condition Monitoring (M. Mezache);
S2/WG13—Shock Test Requirements for Shelf-Mounted and Other Commercial Electronics Systems (B. Lang);

S2/WG39 (S3)—Human Exposure to Mechanical Vibration and Shock (D.D. Reynolds, Chair; R. Dong, Vice Chair).

S2 Inactive Working Group

S2/WG54—Atmospheric Blast Effects (J.W. Reed).

S2 Standards on Mechanical Vibration and Shock

ANSI S2.1-2000 / ISO 2041:1990 American National Standard Vibration and Shock—Vocabulary (Nationally Adopted International Standard)
ANSI S2.2-1959 (R 2006) American National Standard Methods for the Calibration of Shock and Vibration Pickups
ANSI S2.4-1976 (R 2004) American National Standard Method for Specifying the Characteristics of Auxiliary Analog Equipment for Shock and Vibration Measurements
ANSI S2.7-1982 (R 2004) American National Standard Balancing Terminology
ANSI S2.8-2007 American National Standard Technical Information Used for Resilient Mounting Applications
ANSI/ASA S2.9-2008 American National Standard Parameters for Specifying Damping Properties of Materials and System Damping
.ANSI S2.16-1997 (R 2006) American National Standard Vibratory Noise Measurements and Acceptance Criteria of Shipboard Equipment
ANSI S2.17-1980 (R 2004) American National Standard Techniques of Machinery Vibration Measurement
ANSI S2.19-1999 (R 2004) American National Standard Mechanical Vibration—Balance Quality Requirements of Rigid Rotors, Part 1: Determination of Permissible Residual Unbalance, Including Marine Applications
ANSI S2.20-1983 (R 2006) American National Standard Estimating Air Blast Characteristics for Single Point Explosions in Air, with a Guide to Evaluation of Atmospheric Propagation and Effects
ANSI S2.21-1998 (R 2007) American National Standard Method for Preparation of a Standard Material for Dynamic Mechanical Measurements
ANSI S2.22-1998 (R 2007) American National Standard Resonance Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials
ANSI S2.23-1998 (R 2007) American National Standard Single Cantilever Beam Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials
ANSI S2.24-2001 (R 2006) American National Standard Graphical Presentation of the Complex Modulus of Viscoelastic Materials
ANSI S2.25-2004 American National Standard Guide for the Measurement, Reporting, and Evaluation of Hull and Superstructure Vibration in Ships
ANSI S2.26-2001 (R 2006) American National Standard Vibration Testing Requirements and Acceptance Criteria for Shipboard Equipment
ANSI S2.27-2002 (R 2007) American National Standard Guidelines for the Measurement and Evaluation of Vibration of Ship Propulsion Machinery
ANSI S2.28-2003 American National Standard Guide for the Measurement and Evaluation of Vibration of Shipboard Machinery
ANSI S2.29-2003 American National Standard Guide for the Measurement and Evaluation of Vibration of Machine Shafts on Shipboard Machinery
ANSI S2.31-1979 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 1: Basic Definitions and Transducers
ANSI S2.32-1982 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 2: Measurements Using Single-Point Translational Excitation
ANSI S2.34-1984 (R 2005) American National Standard Guide to the Experimental Determination of Rotational Mobility Properties and the Complete Mobility Matrix
ANSI S2.42-1982 (R 2004) American National Standard Procedures for Balancing Flexible Rotors
ANSI S2.43-1984 (R 2005) American National Standard Criteria for Evaluating Flexible Rotor Balance
ANSI S2.46-1989 (R 2005) American National Standard Characteristics to be Specified for Seismic Transducers
ANSI S2.48-1993 (R 2006) American National Standard Servo-Hydraulic Test Equipment for Generating Vibration—Methods of Describing Characteristics
ANSI S2.60-1987 (R 2005) American National Standard Balancing Machines—Enclosures and Other Safety Measures
ANSI S2.61-1989 (R 2005) American National Standard Guide to the Mechanical Mounting of Accelerometers
ANSI S2.70-2006 (*revision of ANSI S3.34-1986*) American National Standard Guide for the Measurement and Evaluation of Human Exposure to Vibration Transmitted to the Hand
ANSI S2.71-1983 (R 2006) (*reaffirmation and redesignation of ANSI S3.29-1983*) American National Standard Guide to the Evaluation of Human Exposure to Vibration in Buildings
ANSI S2.72/Part 1-2002 (R 2007) / ISO 2631-1:1997 (*redesignation of ANSI S3.18/Part 1-2002 / ISO 2631-1:1997*) American National Standard Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 1: General requirements (Nationally Adopted International Standard)
ANSI S2.72/Part 4-2003 (R 2007) / ISO 2631-4:2001 (*redesignation of ANSI S3.18/Part 4-2003 / ISO 2631-4:2001*) American National Standard Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 4: Guidelines for the evaluation of the effects of vibration and rotational motion on passenger and crew comfort in fixed-guideway transport systems (Nationally Adopted International Standard)
ANSI S2.73-2002 / ISO 10819:1996 (R 2007) (*redesignation of ANSI S3.40-2002 / ISO 10819:1996*) American National Standard Mechanical vibration and shock—Hand-arm vibration—Method for the measurement and evaluation of the vibration transmissibility of gloves at the palm of the hand (Nationally Adopted International Standard)

ACCREDITED STANDARDS COMMITTEE ON BIOACOUSTICS, S3

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

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

S3 Working Groups

S3/Advisory—Advisory Planning Committee to S3 (C.A. Champlin, Chair; D.A. Preves, Vice Chair);
S3/WG35—Audiometers (R.L. Grason);
S3/WG36—Speech Intelligibility (R.S. Schlauch);
S3/WG37—Coupler Calibration of Earphones (B. Kruger);
S3/WG39—Human Exposure to Mechanical Vibration and Shock—Parallel to ISO/TC 108/SC 4 (D.D. Reynolds, Chair; R. Dong, Vice Chair);
S3/WG43—Method for Calibration of Bone Conduction Vibrators (J.D. Durrant);
S3/WG48—Hearing Aids (D.A. Preves);
S3/WG51—Auditory Magnitudes (R.P. Hellman);
S3/WG56—Criteria for Background Noise for Audiometric Testing (J. Franks);
S3/WG59—Measurement of Speech Levels (M.C. Killion and L.A. Wilber, Co-Chairs);
S3/WG60—Measurement of Acoustic Impedance and Admittance of the Ear (Vacant);
S3/WG62—Impulse Noise with Respect to Hearing Hazard (J.H. Patterson, Chair; R. Hamernik, Vice Chair);
S3/WG67—Manikins (M.D. Burkhard);
S3/WG72—Measurement of Auditory Evoked Potentials (R.F. Burkard);
S3/WG76—Computerized Audiometry (A.J. Miltich);
S3/WG79—Methods for Calculation of the Speech Intelligibility Index (C.V. Pavlovic);
S3/WG81—Hearing Assistance Technologies (L. Thibodeau and L.A. Wilber, Co-Chairs);
S3/WG82—Basic Vestibular Function Test Battery (C. Wall);
S3/WG83—Sound Field Audiometry (T.R. Letowski);
S3/WG84—Otoacoustic Emissions (G. R. Long);
S3/WG88—Standard Audible Emergency Evacuation and Other Signals (Vacant);

S3/WG89—Spatial Audiometry in Real and Virtual Environments (J. Besing);
S3/WG91—Text-to-Speech Synthesis Systems (C. Bickley and A. K. Syrdal, Co-Chairs).

S3 Liaison Group

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

S3 Inactive Working Groups

S3/WG71—Artificial Mouths (R. L. McKinley);
S3/WG80—Probe-tube Measurements of Hearing Aid Performance (W. A. Cole);
S3/WG85—Allocation of Noise-induced Hearing Loss (R. A. Dobie).

S3 Standards on Bioacoustics

ANSI S3.1-1999 (R 2003) American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms
ANSI S3.2-1989 (R 1999) American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems
ANSI S3.4-2007 American National Standard Procedure for the Computation of Loudness of Steady Sounds
ANSI S3.5-1997 (R 2007) American National Standard Methods for Calculation of the Speech Intelligibility Index
ANSI S3.6-2004 American National Standard Specification for Audiometers
ANSI S3.7-1995 (R 2003) American National Standard Method for Coupler Calibration of Earphones
ANSI S3.13-1987 (R 2007) American National Standard Mechanical Coupler for Measurement of Bone Vibrators
ANSI S3.20-1995 (R 2003) American National Standard Bioacoustical Terminology
ANSI S3.21-2004 American National Standard Methods for Manual Pure-Tone Threshold Audiometry
ANSI S3.22-2003 American National Standard Specification of Hearing Aid Characteristics
ANSI S3.25-1989 (R 2003) American National Standard for an Occluded Ear Simulator
ANSI S3.35-2004 American National Standard Method of Measurement of Performance Characteristics of Hearing Aids under Simulated Real-Ear Working Conditions
ANSI S3.36-1985 (R 2006) American National Standard Specification for a Manikin for Simulated *in situ* Airborne Acoustic Measurements
ANSI S3.37-1987 (R 2007) American National Standard Preferred Earhook Nozzle Thread for Postauricular Hearing Aids
ANSI S3.39-1987 (R 2007) American National Standard Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance)
ANSI S3.41-1990 (R 2001) American National Standard Audible Emergency Evacuation Signal
ANSI S3.42-1992 (R 2007) American National Standard Testing Hearing Aids with a Broad-Band Noise Signal
ANSI S3.44-1996 (R 2006) American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment
ANSI S3.45-1999 American National Standard Procedures for Testing Basic Vestibular Function
ANSI S3.46-1997 (R 2007) American National Standard Methods of Measurement of Real-Ear Performance Characteristics of Hearing Aids

ACCREDITED STANDARDS COMMITTEE ON ANIMAL BIOACOUSTICS, S3/SC 1

(D.K. Delaney, Chair; M. C. Hastings, Vice Chair)

Scope: 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.

S3/SC 1 Working Groups

S3/SC1/WG1—Animal Bioacoustics Terminology (A. E. Bowles);
S3/SC1/WG2—Effects of Sound on Fish and Turtles (R. R. Fay and A. N. Popper, Co-Chairs);
S3/SC 1/WG3—Underwater Passive Acoustic Monitoring for Bioacoustic Applications (A. M. Thode).

ACCREDITED STANDARDS COMMITTEE ON NOISE, S12

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

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

S12 Working Groups

S12/Advisory—Advisory Planning Committee to S12 (R. D. Hellweg; W. J. Murphy, Vice Chair);
S12/WG3—Measurement of Noise from Information Technology and Telecommunications Equipment (K. X. C. Man);
S12/WG11—Hearing Protector Attenuation and Performance (E. H. Berger);
S12/WG13—Method for the Selection of Hearing Protectors that Optimize the Ability to Communicate (D. Byrne);
S12/WG14—Measurement of the Noise Attenuation of Active and/or Passive Level Dependent Hearing Protective Devices (W. J. Murphy);
S12/WG15—Measurement and Evaluation of Outdoor Community Noise (P. D. Schomer);
S12/WG18—Criteria for Room Noise (R. J. Peppin);
S12/WG23—Determination of Sound Power (B. M. Brooks and J. Schmitt, Co-chairs);
S12/WG31—Predicting Sound Pressure Levels Outdoors (L. Pater);
S12/WG32—Revision of ANSI S12.7-1986 Methods for Measurement of Impulse Noise (W. Ahroon);
S12/WG36—Development of Methods for Using Sound Quality (P. Davies and G. L. Ebbitt, Co-Chairs);
S12/WG38—Noise Labeling in Products (R. D. Hellweg);
S12/WG40—Measurement of the Noise Aboard Ships (S. Antonides, Chair; S.A. Fisher, Vice Chair);
S12/WG41—Model Community Noise Ordinances (L. S. Finegold, Chair; B. M. Brooks, Vice Chair);
S12/WG44—Speech Privacy (G. C. Tocci, Chair; D. Sykes, Vice Chair);
S12/WG45—Measurement of Occupational Noise Exposure from Telephone Equipment (K. A. Woo, Chair; L. A. Wilber, Vice Chair);
S12/WG46—Acoustical Performance Criteria for Relocatable Classrooms (T. Hardiman and P.D. Schomer, Co-Chairs);
S12/WG47—Underwater Noise Measurements of Ships (M. Bahtiarian);
S12/WG48—Railroad Horn Sound Emission Testing (J. Erdreich, Chair; J.J. Earshen, Vice-Chair);
S12/WG49—Noise from Hand-operated Power Tools, Excluding Pneumatic Tools (B.M. Brooks);
S12/WG50—Information Technology (IT) Equipment in Classrooms (R.D. Hellweg);
S12/WG51—Procedure for Measuring the Ambient Noise Level in a Room (J.G. Lilly).

S12 Liaison Groups

S12/L-1 IEEE 85 Committee for TAG Liaison—Noise Emitted by Rotating Electrical Machines (Parallel to ISO/TC43/SC 1/WG 13) (R. G. Bartheld);
S12/L-2 Measurement of Noise from Pneumatic Compressors Tools and Machines (Parallel to ISO/TC43/SC 1/WG 9) (Vacant);
S12/L-3 SAE Committee for TAG Liaison on Measurement and Evaluation of Motor Vehicle Noise (parallel to ISO/TC 43/SC 1/WG 8) (R. F. Schumacher);
S12/L-4 SAE Committee A-21 for TAG Liaison on Measurement and Evaluation of Aircraft Noise (J. D. Brooks);
S12/L-5 ASTM E-33 on Environmental Acoustics (to include activities of ASTM E33.06 on Building Acoustics, parallel to ISO/TC43/SC 2 and ASTM E33.09 on Community Noise) (K. P. Roy);
S12/L-6 SAE Construction-Agricultural Sound Level Committee (I. Douell);
S12/L-7 SAE Specialized Vehicle and Equipment Sound Level Committee (T.M. Disch);
S12/L-8 ASTM PTC 36 Measurement of Industrial Sound (R.A. Putnam, Chair; B.M. Brooks, Vice Chair).

S12 Inactive Working Groups

S12/WG9—Annoyance Response to Impulsive Noise (L. C. Sutherland);
S12/WG19—Measurement of Occupational Noise Exposure (J. Barry);
S12/WG27—Outdoor Measurement of Sound Pressure Level (G. A. Daigle);
S12/WG29—Field Measurement of the Sound Output of Audible Public-Warning Devices (Sirens) (P. Graham);
S12/WG37—Measuring Sleep Disturbance Due to Noise (K.S. Pearsons).

S12 Standards on Noise

ANSI S12.1-1983 (R 2006) American National Standard Guidelines for the Preparation of Standard Procedures to Determine the Noise Emission from Sources
ANSI/ASAS12.2-2008 American National Standard Criteria for Evaluating Room Noise
ANSI S12.3-1985 (R 2006) American National Standard Statistical Methods for Determining and Verifying Stated Noise Emission Values of Machinery and Equipment
ANSI S12.5-2006/ISO 6926:1999 American National Standard Acoustics—Requirements for the Performance and Calibration of Reference Sound Sources Used for the Determination of Sound Power Levels (Nationally Adopted International Standard)
ANSI/ASA S12.6-2008 American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors
ANSI S12.7-1986 (R 2006) American National Standard Methods for Measurements of Impulse Noise
ANSI/ASAS12.8-1998 (R 2008) American National Standard Methods for Determining the Insertion Loss of Outdoor Noise Barriers
ANSI S12.9/Part 1-1988 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 1
ANSI/ASAS12.9/Part 2-1992 (R 2008) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 2: Measurement of Long-Term, Wide-Area Sound
ANSI/ASA S12.9/Part 3-1993 (R 2008) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 3: Short-Term Measurements with an Observer Present
ANSI S12.9/Part 4-2005 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 4: Noise Assessment and Prediction of Long-Term Community Response
ANSI/ASA S12.9/Part 5-2007 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 5: Sound Level Descriptors for Determination of Compatible Land Use
ANSI/ASA S12.9/Part 6-2008 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 6: Methods for Estimation of Awakenings Associated with Outdoor Noise Events Heard in Homes
ANSI/ASA S12.10-2002 (R 2007)/ISO 7779:1999 American National Standard Acoustics—Measurement of airborne noise emitted by information technology and telecommunications equipment (Nationally Adopted International Standard)
ANSI/ASA S12.11/Part 1-2003/ISO 10302:1996 (MOD) (R 2008) American National Standard Acoustics—Measurement of noise and vibration of small air-moving devices—Part 1: Airborne noise emission (Modified Nationally Adopted International Standard)
ANSI/ASA S12.11/Part 2—2003 (R 2008) American National Standard Acoustics—Measurement of Noise and Vibration of Small Air-Moving Devices—Part 2: Structure-Borne Vibration
ANSI/ASA S12.12-1992 (R 2007) American National Standard Engineering Method for the Determination of Sound Power Levels of Noise Sources Using Sound Intensity
ANSI S12.13 TR-2002 ANSI Technical Report Evaluating the Effectiveness of Hearing Conservation Programs through Audiometric Data Base Analysis
ANSI/ASA S12.14-1992 (R 2007) American National Standard Methods for the Field Measurement of the Sound Output of Audible Public Warning Devices Installed at Fixed Locations Outdoors
ANSI/ASA S12.15-1992 (R 2007) American National Standard For Acoustics—Portable Electric Power Tools, Stationary and Fixed Electric Power Tools, and Gardening Appliances—Measurement of Sound Emitted
ANSI/ASA S12.16-1992 (R 2007) American National Standard Guidelines for the Specification of Noise of New Machinery
ANSI S12.17-1996 (R 2006) American National Standard Impulse Sound Propagation for Environmental Noise Assessment
ANSI S12.18-1994 (R 2004) American National Standard Procedures for Outdoor Measurement of Sound Pressure Level
ANSI S12.19-1996 (R 2006) American National Standard Measurement of Occupational Noise Exposure
ANSI S12.23-1989 (R 2006) American National Standard Method for the Designation of Sound Power Emitted by Machinery and Equipment
ANSI S12.42-1995 (R 2004) American National Standard Microphone-in-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Circumaural Hearing Protection Devices
ANSI/ASA S12.43-1997 (R 2007) American National Standard Methods for Measurement of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions
ANSI/ASA S12.44-1997 (R 2007) American National Standard Methods for Calculation of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions from Sound Power Level
ANSI/ASA S12.50-2002 (R 2007)/ISO 3740:2000 American National Standard Acoustics—Determination of sound power levels of noise sources—Guidelines for the use of basic standards (Nationally Adopted International Standard)
ANSI/ASA S12.51-2002 (R 2007)/ISO 3741:1999 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision method for reverberation rooms (This Standard includes Technical Corrigendum 1-2001.) (Nationally Adopted International Standard)
ANSI S12.53/Part 1-1999 (R 2004)/ISO 3743-1:1994 American National Standard Acoustics—Determination of sound power levels of noise sources—Engineering methods for small, movable sources in reverberant fields—Part 1: Comparison method for hard-walled test rooms (Nationally Adopted International Standard)
ANSI S12.53/Part 2-1999 (R 2004)/ISO 3743-2:1994 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering methods for small, movable sources in reverberant fields—Part 2: Methods for special reverberation test rooms (Nationally Adopted International Standard)
ANSI S12.54-1999 (R 2004)/ISO 3744:1994 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering method in an essentially free field over a reflecting plane (Nationally Adopted International Standard)
ANSI S12.55-2006/ISO 3745:2003 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for anechoic and hemi-anechoic rooms (Nationally Adopted International Standard)
ANSI S12.56-1999 (R 2004)/ISO 3746:1995 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Survey method using an enveloping measurement surface over a reflecting plane (Nationally Adopted International Standard)
ANSI/ASA S12.57-2002 (R 2007)/ISO 3747:2000 American National Stan-

dard Acoustics—Determination of sound power levels of noise sources using sound pressure—Comparison method *in situ* (Nationally Adopted International Standard)
ANSI S12.60-2002 American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools
ANSI S12.65-2006 (*Revision of ANSI S3.14-1977*) American National Standard for Rating Noise with Respect to Speech Interference
ANSI/ASA S12.68-2007 American National Standard Methods of Estimating Effective A-Weighted Sound Pressure Levels When Hearing Protectors are Worn

ASA COMMITTEE ON STANDARDS (ASACOS)

ASACOS (P.D. Schomer, Chair; R.D. Hellweg, Vice Chair)

U.S. TECHNICAL ADVISORY GROUPS (TAGS) FOR INTERNATIONAL STANDARDS COMMITTEES

ISO/TC 43 Acoustics, **ISO/TC 43/SC 1** Noise (P.D. Schomer, U.S. TAG Chair)
ISO/TC 108 Mechanical vibration, shock, and condition monitoring (D.J. Evans, U.S. TAG Chair)
ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures (A.F. Kilcullen and R.F. Taddeo, U.S. TAG Co-Chairs)
ISO/TC 108/SC 3 Use and calibration of vibration and shock measuring instruments (D.J. Evans, U.S. TAG Chair)
ISO/TC 108/SC 4 Human exposure to mechanical vibration and shock (D.D. Reynolds, U.S. TAG Chair)
ISO/TC 108/SC 5 Condition monitoring and diagnostics of machines (D.J. Vendittis, U.S. TAG Chair; R. Taddeo, U.S. TAG Vice Chair)
ISO/TC 108/SC 6 Vibration and shock generating systems (C. Peterson, U.S. TAG Chair)
IEC/TC 29 Electroacoustics (V. Nedzelnitsky, U.S. Technical Advisor)

STANDARDS NEWS FROM THE UNITED STATES

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

American National Standards Call for Comment on Proposals Listed

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

AHRI (Air-Conditioning, Heating, and Refrigeration Institute)

New Standards

BSR/ARI 1140-200x, Sound Quality Evaluation Procedures for Air-Conditioning and Refrigeration Equipment (new standard)

Provides guidance of the industry, including manufacturers, engineers installers, contractors, and users. (September 29, 2008)

ASA (ASC S3) (Acoustical Society of America)

Reaffirmations

BSR/ASA S3.1-1999 (R200x), Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms (reaffirmation and redesignation of ANSI S3.1-1999 (R2003))

Specifies maximum permissible ambient noise levels (MPANLs) allowed in an audiometric test room that produce negligible masking (<2 dB) of test signals presented at reference equivalent threshold levels specified in ANSI S3.6-1996. The MPANLs are specified from 125 to 8000 Hz in octave and one-third octave band intervals for two audiometric testing conditions (ears covered and ears not covered) and for three test frequency ranges (125 to 8000 Hz, 250 to 8000 Hz, and 500 to 8000 Hz). (October 6, 2008)

BSR/ASA S3.7-1995 (R200x), Method for Coupler Calibration of Earphones (reaffirmation and redesignation of ANSI S3.7-1995 (R2003))

Specifies the physical configuration and acoustical performance of couplers for calibration of supra-aural and insert earphones are specified. This standard, which is a revision of ANSI S3.7-1973, describes a family of 6.0 cm^3 and 2.0 cm^3 couplers, and provides information on the methods for coupler calibration of the respective supra-aural and insert earphones. (October 6, 2008)

BSR/ASA S3.20-1995 (R200x), Bioacoustical Terminology (reaffirmation and redesignation of ANSI S3.20-1995 (R2003))

Provides definitions for a wide variety of terms used in human bioacoustics, including hearing, speech, psychoacoustics, and physiological acoustics. It is intended to supplement ANSI S1.1-1994 (R2004), American National Standard Acoustical Terminology, in which more-generally used terms in acoustics are defined, including a number of terms from physiological and psychological acoustics and music. Terms from ANSI S1.1-1994 that are related to bioacoustics are included in this standard as annexes. (October 6, 2008)

TIA (Telecommunications Industry Association)

Reaffirmations

BSR/TIA 825-A-2003 (R200x), A Frequency Shift Keyed Modem for Use on the Public Switched Telephone Network (reaffirmation of ANSI/TIA 825-A-2003)

Specifies a Frequency Shift Keyed (FSK) modem operating at a data signaling rates of 50 and 45.45 symbols/sec. The deaf and hard-of-hearing generally use this modem for real time 2-way text-based communication over the public switched telephone network.

UL (Underwriters Laboratories, Inc.)

Revisions

BSR/UL 217-200x, Single and Multiple Station Smoke Alarms (revision of ANSI/UL 217-2006) Revises proposals for voice message in alarm signal and removal of inaccurate information for sound output measurement. (August 17, 2008)

Call for Members (ANS Consensus Bodies)

Directly and materially affected parties who are interested in participating as a member of an ANS consensus body for the standards listed below are requested to contact the sponsoring standards developer directly and in a timely manner.

AHRI (Air-Conditioning, Heating, and Refrigeration Institute)

BSR/ARI 1140-200x, Sound Quality Evaluation Procedures for Air-Conditioning and Refrigeration Equipment (new standard)

ASA (ASC S1) (Acoustical Society of America)

BSR/ASA S1.17-200X/Part 2-200x, Microphone Windscreens—Part 2: Measurement of Windscreen Noise in Moving Air (new standard)

ASA (ASC S3) (Acoustical Society of America)

BSR/ASA S3.6-200x, Specification for Audiometers (revision and redesignation of ANSI S3.6-2004)
BSR/ASA S3.22-200x, Specification of Hearing Aid Characteristics (revision and redesignation of ANSI S3.22-2003)
BSR/ASA S3.51-200x, Procedure for the generation and measurement of acoustic stimuli used to elicit auditory evoked potentials (new standard)

ASA (ASC S12) (Acoustical Society of America)

BSR/ASA S12.72-200x, Procedure for Measuring the Ambient Noise Level in a Room (new standard)

Project Initiation Notification System (PINS)

ANSI Procedures require notification of ANSI by ANSI-accredited standards developers of the initiation and scope of activities expected to result in new or revised American National Standards. This information is a key element in planning and coordinating American National Standards. The following is a list of proposed new American National Standards or revisions to existing American National Standards that have been received from ANSI-accredited standards developers that utilize the periodic maintenance option in connection with their standards. Directly and materially affected interests wishing to receive more information should contact the standards developer directly.

ASA (ASC S1) (Acoustical Society of America)

BSR/ASA S1.17-200X/Part 2-200x, Microphone Windscreens—Part 2: Measurement of Windscreen Noise in Moving Air (new standard)

Specifies a test to measure the wind-generated measurement noise created by the windscreen via its interaction with moving air over a defined audio frequency range. Project Need: To address the condition where moving laminar or turbulent air flow over the windscreen creates noise that reaches the windscreened microphone. Stakeholders: Acoustical engineers and all users of sound measurement equipment.

ASA (ASC S3) (Acoustical Society of America)

BSR/ASA S3.6-200x, Specification for Audiometers (revision and redesignation of ANSI S3.6-2004)

Provides specifications and tolerances for pure-tone, speech, and masking signals and describes the minimum test capabilities of different types of audiometers. The audiometers covered in this specification are devices designed for use in determining the hearing threshold level of an individual in comparison with a chosen standard reference threshold level. Project Need: To correct and update the 2004 edition of this standard and incorporate changes to the revised IEC and ISO equivalent or referenced standards in order to maintain technical equivalence. Stakeholders: Audiometer manufacturers, audiologists, calibration personal, certification/testing laboratories.

BSR/ASA S3.22-200x, Specification of Hearing Aid Characteristics (revision and redesignation of ANSI S3.22-2003)

Describes air-conduction hearing-aid measurement methods that are suitable for specification and tolerance purposes. Various test methods are described. Configurations are given for measuring the input SPL to a hearing aid. Tolerances in relation to values specified by the manufacture are given for certain parameters. Appendices describe equivalent substitution methods, characteristics of battery simulators, and additional tests to characterize the electroacoustic performance of hearing aids. Project Need: To update the 2003 edition to follow the existing practice of using a linear test mode. Stakeholders: Hearing aid manufacturers, hearing aid dispensers.

BSR/ASA S3.51-200x, Procedure for the generation and measurement of acoustic stimuli used to elicit auditory evoked potentials (new standard)

Identifies minimum specifications for the stimuli used to obtain auditory evoked potentials (including otoacoustic emissions), as well as recommended calibration procedures. Project Need: To address calibration procedures and instrumentation specifications for the stimuli typically used for obtaining auditory evoked potentials (including otoacoustic emissions), such as transients. This standard will provide information that is missing from ANSI/ASA S3.6. Stakeholders: Audiologists, acousticians, and businesses that produce auditory-evoked potential instrumentation.

ASA (ASC S12) (Acoustical Society of America)

BSR/ASA S12.70-200x, Criteria for Evaluating Speech Privacy in Healthcare Facilities (new standard)

Includes normative references, definitions, speech privacy descriptor definition, method for measuring the speech privacy descriptor, and recommended descriptor values based on adjacent space uses. Project Need: To enforce Speech Privacy protection statutes by providing a practical, measurable standard of enforcement and active, knowledgeable participation by acoustical professionals. Stakeholders: Architects, acoustical consultants, planners, engineers and designers of healthcare facilities.

BSR/ASA S12.72-200x, Procedure for Measuring the Ambient Noise Level in a Room (new standard)

The standard will specify how to conduct ambient noise measurements in a room. Specifically, it will address instruction requirements, microphone placement, duration of each measurement, identification of specific noise sources (e.g., HVAC, lighting, electrical, exterior traffic and aircraft) and data reduction and analysis. Project Need: To mitigate the ambient noise level in a room, which may vary significantly with both microphone location and time. In addition, other factors (e.g., HVAC operating conditions) can significantly alter the measured sound values. Stakeholders: Architects, acoustical consultants, building and construction engineers.

ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.)

BSR/ASHRAE Standard 197-200x, Method of Test for Attenuation Characteristics of Vibration Isolators (new standard)

Vibration isolators commonly used to prevent the transmission of HVACR equipment vibration into buildings, such as springs, fiberglass mounts, elastomeric mounts and pads, wire rope or cork pads. Project Need: To prescribe the method of test to measure the performance of isolators over a broad range of frequencies. Stakeholders: Building owners, MEP engineers, structural engineers, acoustical (noise and vibration) consultants.

ASME (American Society of Mechanical Engineers)

BSR/ASME PTC 36-200x, Measurement of Industrial Sound (revision of ANSI/ASME PTC 36-2004)

Describes procedures for measuring and reporting airborne sound emission from stationary sound sources and equipment, or from facilities composed of multiple stationary sound sources. Project Need: To provide a periodic review of the Code as well as to consider the inclusion of relevant portions of ASME B133.8 Standard. Stakeholders: Power plant personnel and employees at other installations.

TIA (Telecommunications Industry Association)

BSR/TIA 1083-A-200x, Telecommunications—Telephone Terminal Equipment—Handset—Magnetic—Measurement Procedures and Performance Requirements (revision of ANSI/TIA 1083-2007)

Defines measurement procedures and performance requirements for the handset-generated audio band magnetic noise of wireline telephones. Project Need: To cover a consumer product or service. Stakeholders: Telecommunications Industry Association.

Final Actions on American National Standards

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

ASA (ASC S12) (Acoustical Society of America)

Revisions

ANSI/ASA S12.2-2008, Criteria for Evaluating Room Noise (revision and redesignation of ANSI S12.2-1995 (R1999))

ANSI/ASA S12.9-Part 6-2008, Quantities and Procedures for Description and Measurement of Environmental Sound—Part 6: Methods for Estimating of Awakenings Associated with Outdoor Noise Events Heard in Homes (revision and redesignation of ANSI S12.9-Part 6-2000 (R2005))

Reaffirmations

ANSI/ASA S12.11-2003 Part 1/ISO 10302:1996 (MOD) (R2008), Acoustics—Measurement of noise and vibration of small air-moving devices, Part 1: Airborne noise emission (a modified NAIS) (reaffirmation and redesignation of ANSI S12.11/1 ISO 10302-1996 (MOD)-2003)

ANSI/ASA S12.11-2003/Part 2 (R2008), Acoustics—Measurement of noise and vibration of small air-moving devices, Part 2: Structure-borne vibration (reaffirmation and redesignation of ANSI S12.11/2-2003)

IEEE (Institute of Electrical and Electronics Engineers)

Reaffirmations

ANSI/IEEE 260.4-1996 (R2008), Letter Symbols and Abbreviations for Quantities Used in Acoustics (reaffirmation of ANSI/IEEE 260.4-1996 (R2002))

STANDARDS NEWS FROM ABROAD

Newly Published ISO and IEC Standards

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

IEC Standards

PERFORMANCE OF HOUSEHOLD ELECTRICAL APPLIANCES (TC59)

IEC 60704-2-13 Ed. 1.2 b:2008, Household and similar electrical appliances—Test code for the determination of airborne acoustical noise—Part 2-13: Particular requirements for range hoods

IEC/PAS 60704-2-15 Ed. 1.0 en:2008, Household and similar electrical appliances—Test code for the determination of airborne acoustical noise—Part 2-15: Particular requirements for household food waste disposers

FIBRE OPTICS (TC 86)

IEC 61290-3-2 Ed. 2.0 b:2008, Optical amplifiers—Test methods—Part 3-2: Noise figure parameters—Electrical spectrum analyzer method

ISO Draft Standards

HYDROMETRIC DETERMINATIONS (TC 113)

ISO/DIS 15769, Hydrometry—Guidelines for the application of acoustic velocity meters using the doppler and echo correlation methods (11/16/2008)

ADVANCED-DEGREE DISSERTATIONS IN ACOUSTICS

Editor's Note: Abstracts of Doctoral and Master's theses will be welcomed at all times. Please note that they must be limited to 200 words, must include the appropriate PACS classification numbers, and must be formatted as shown below. If sent by postal mail, note that they must be double spaced. The address for obtaining a copy of the thesis is helpful. Submit abstracts to: Acoustical Society of America, Thesis Abstracts, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, e-mail: asa@aip.org

Geometric and Boundary Element Method Simulations of Acoustic Reflections From Rough, Finite, or Nonplanar Surfaces [43.55.Ka, 43.55.Br] —Jonathan Rathsam, *Charles W. School of Architectural Engineering and Construction, University of Nebraska-Lincoln, Omaha, NE, May 2008 (Ph.D.).* This dissertation seeks to advance the current state of computer-based sound field simulations for room acoustics. Acoustical designers commonly use *geometric* simulations for quick sound field predictions. A geometric simulation of reflections from rough surfaces is still under refinement. The first project in this dissertation investigates the scattering coefficient, which quantifies the degree of diffuse reflection from rough surfaces. The main result is that predicted reverberation time varies inversely with scattering coefficient *if* the sound field is nondiffuse. Geometric acoustics is a high-frequency approximation to wave acoustics. Acoustical designers encounter the limits of geometric acoustics when simulating the low-frequency response from finite suspended reflector panels. The second project in this dissertation uses the rigorous boundary element method (BEM) to develop an improved low-frequency radiation model for smooth finite reflectors. The improved low-frequency model is suggested in two forms for implementation in geometric simulations. The final project in this dissertation uses BEM to investigate the sound field around nonplanar reflectors. The author has added convex edges rounded away from the source side of a finite smooth reflector to minimize coloration of reflections caused by boundary-wave interference. [Copies may be obtained from Jonathan Rathsam, jrathsam@gmail.com]

Thesis Advisor: Lily M. Wang

Investigations of Incorporating Source Directivity Into Room Acoustics Computer Models to Improve Auralizations [43.55.Ka, 43.55.Hy] —Michelle C. Vigeant, *Architectural Engineering Program, University of Nebraska-Lincoln, Omaha, NE, August 2008 (Ph.D.).* Room acoustics computer modeling and auralizations are useful tools when designing or modifying acoustically sensitive spaces. In this dissertation, the input parameter of source directivity has been studied in great detail to determine, first, its effect in room acoustics computer models and, second, how to better incorporate the directional source characteristics to improve auralizations. The room acoustics computer model was initially validated in terms of accurately incorporating the input source directivity. The next study demonstrated that using directional sources over an omnidirectional source in room acoustics computer models produces significant differences both in terms of calculated room acoustics parameters and auralizations. A recently proposed technique for creating auralizations using multichannel anechoic recordings has been examined with numerous subjective studies, applied to both solo instruments and an orchestra. Through these studies, this process was shown to be effective in terms of improving the realism and source width of the auralizations in a number of cases and also modeling different source orientations. In addition, this approach was applied to modeling an entire orchestra with individual sources, in three different configurations, for the first time. This method shows great promise as a new technique for auralizing both solo instruments and an entire orchestra. [Copies may be obtained from Michelle Vigeant, michelle.vigeant@gmail.com]

Thesis advisor: Lily M. Wang

 © 2008 Acoustical Society of America

BOOK REVIEWS

Brian C. J. Moore
John Wiley & Sons, New Jersey, 2007

Cochlear Hearing Loss: Physiological, Psychological and Technical Issues, Second Edition (Wiley Series in Human Communication Science)

Brian C. J. Moore

John Wiley & Sons, Chichester, 2007 332 pp. Price: $90.00 (paperback), ISBN-13: 978-0470516331

In the preface to *Cochlear Hearing Loss*, Brian Moore notes that a primary aim of the book is "to convey an understanding of the perceptual changes associated with cochlear hearing loss, of the difficulties faced by the hearing-impaired person and the possibilities and limitations of current hearing aids. The reader should come away with an impression not only of what happens but also of why it happens. I have attempted to write at a level that is accessible and comprehensible to students in audiology and speech science. I hope, therefore, that the book will be considered suitable as textbook for such students." The book largely succeeds in meeting these goals.

The book is divided into nine chapters. The organization of the chapters is to first provide a foundation in the functioning of the normal auditory system followed by a review of how a damaged cochlea affects this functioning. Each chapter then concludes with a discussion of how specific consequences of cochlear hearing loss affect auditory perception. Throughout the book, the perceptual effects of cochlear hearing loss are linked to underlying mechanisms, including the physiology of the damaged cochlea. As a second edition (originally published in 1998), this book has been significantly updated and revised to reflect recent advances in hearing science and clinical audiology. Readers who have studied Brian Moore's other books will find many of the figures in the text familiar.

The first chapter begins with an introduction to the concepts of linear and nonlinear systems, which are essential ideas in understanding the functioning of the normal auditory system as well as the consequences of cochlear damage. The first chapter then presents an overview of the basic structure and function of the normal peripheral auditory system, beginning with the conductive mechanism and then moving to the cochlea and auditory nerve. This overview includes a discussion of basilar-membrane tuning, the active and passive mechanisms within the cochlea, the functional roles of inner and outer hair cells, and the cochlear nonlinearity and its manifestations (e.g., two-tone suppression, combination tones, otoacoustic emissions). The overview also addresses the nature of neural responses to sound in the mammalian auditory nerve. Chapter 1 then delves into how cochlear and neural responses to sound change due to outer and/or inner hair cell damage. This chapter leads readers to an understanding of how the active mechanism and the compressive nonlinearity of the cochlea are altered due to outer hair cell loss.

Chapter 2 provides a succinct discussion of auditory sensitivity, including different ways to quantify absolute thresholds (through earphones vs in sound field), the clinical audiogram which shows how sensitivity changes as function of frequency, and how cochlear hearing loss can cause a loss in auditory sensitivity. This chapter also introduces the idea of cochlear dead regions, which are regions of the damaged cochlea with minimal or no inner hair cell function. A key concept emerging from this chapter is that cochlear hearing loss can cause changes to sensitivity (threshold effects) as well as changes to the processing of sounds that are above a person's absolute thresholds (suprathreshold effects). Most of the remaining chapters address suprathreshold processing changes that result from cochlear damage.

Chapter 3, "Masking, Frequency Selectivity and Basilar Membrane Nonlinearity," discusses the frequency resolution of the auditory system and how frequency resolution is altered when the nonlinear mechanisms of the cochlea are damaged. The chapter systematically describes different ways that masking is used to measure the frequency resolution in both normal and impaired auditory systems. Masking procedures include psychophysical tuning curves, notched-noise methods, masking patterns, estimation of excitation patterns, and methods using both simultaneous and nonsimultaneous masking. The chapter provides an overview of the characteristics of auditory filters in both normal and impaired systems, and also addresses how degraded frequency resolution affects auditory and speech perception in persons with cochlear hearing loss.

Chapter 4 focuses on the perception of loudness and the ability of human listeners to resolve changes in intensity. The chapter begins with descriptions of how loudness perception is measured in listeners with normal hearing including equal-loudness contours, scales of loudness, and the detection of changes in intensity. The chapter then addresses how loudness perception and intensity resolution are affected by cochlear hearing loss, including the clinically important concept of loudness recruitment. The chapter also provides in-depth descriptions of the author's models of loudness perception based on both systems with normal auditory function and systems with cochlear damage.

Chapter 5 presents an overview of temporal resolution and temporal integration in the normal auditory system, and then describes evidence regarding why listeners with cochlear hearing loss have degraded temporal resolution. Four stages of a model of temporal resolution in the normal auditory system are first presented (bandpass filtering, compressive nonlinearity, sliding temporal integrator, and a decision device). Then, data from studies of people with normal hearing and studies of people with cochlear hearing loss are discussed and analyzed in the context of the temporal resolution models. The implications of temporal processing deficits on speech perception are also discussed.

The topic of the sixth chapter is pitch perception and frequency discrimination. This chapter provides an overview of theoretical and experimental studies of both pure-tone and complex-tone perception. Pitch perception in persons with normal hearing is considered first, followed by a discussion of how cochlear hearing loss affects the ability to discriminate frequency and perceive the pitch of complex tones. This chapter builds on many of the concepts introduced in previous chapters, since deficits in both frequency and temporal resolution have been implicated in the pitch perception deficits observed in persons with cochlear hearing loss. The chapter conveys to the reader the importance of pitch in the perception of speech and music, and how cochlear hearing loss can adversely affect this component of complex perception.

Chapter 7 considers the effects of cochlear hearing loss on spatial hearing and the advantages persons receive from listening binaurally. Two specific aspects of localization are addressed: (1) the perceived sound direction relative to the actual sound direction, and (2) the ability of listeners to perceive a small change in the location of a sound source. The chapter begins with a discussion of sound source localization and lateralization for both simple and complex sounds first in persons with normal hearing and then in persons with cochlear hearing loss. The chapter also delves into a number of issues related to binaural hearing and cochlear hearing loss including the role of head movements and the head shadow, the precedence effect, binaural masking level differences, and release from information masking.

Bringing together many of the concepts introduced earlier in the book, Chap. 8 addresses two primary issues that may contribute to degraded speech perception in listeners with cochlear hearing loss: audibility (loss of sensitivity) and suprathreshold processing deficits. In its discussion of audibility, the chapter considers the speech intelligibility index, speech perception at high sound levels in both quiet and in noise, and simulations of hearing loss. The chapter discusses the effects of cochlear dead regions on speech perception and provides a comprehensive review of studies that correlate the relationship between basic psychoacoustic abilities and speech perception. The discussion highlights recent work on the effects of cochlear hearing loss in the processing of temporal fine structure. The chapter also reviews studies that simulate cochlear hearing loss in an effort to delineate the specific psychoacoustic factors that may underlie degraded speech perception.

The final chapter addresses the benefits and limitations of hearing aids, procedures for hearing-aid fittings, and future directions for hearing-aid re-

search and development. The chapter begins with a discussion of linear amplification and prescriptive methods aimed at restoring audibility of the speech signal. The chapter provides a comprehensive look at compression amplification, including basic concepts and definitions as well as a discussion of strategies for fitting compression systems to individual listeners. The author also delineates limitations of current hearing aids and discusses newly developed signal processing strategies for hearing aid applications.

The book concludes with a comprehensive glossary and reference list. The glossary defines terms used in this book as well as in the general hearing science literature. The reference list includes a wide scope of literature related to cochlear hearing loss and includes citations in physiology, basic hearing sciences, and clinically applied research. A significant number of new references have been added since the first edition.

In summary, this book provides readers with a foundation in both "how" and "why" cochlear hearing loss adversely affects auditory perception, including the perception of speech in noisy environments. *Cochlear Hearing Loss* serves as a important resource for students and professionals in hearing-aid engineering, otology, audiology, and the hearing sciences. The book will also serve as an excellent text in psychoacoustics-hearing science courses for graduate students in audiology, especially students in AUD programs. I have taught from the first edition of this text for several years. The book has worked well as a textbook, but has required supplementation with additional background resources for students. I anticipate that the second edition will continue to be an appropriate foundation textbook for graduate students in audiology. The new edition has a notable emphasis on the science underlying several recent developments in clinical audiology, including cochlear dead regions, compression, loudness models and their use in hearing-aid fitting methodologies, the assessment of speech perception in noisy backgrounds, and the limitations and potential of state-of-the-art hearing aids.

KATHRYN H. AREHART
Speech, Language, Hearing Sciences
409 UCB
University of Colorado
Boulder, Colorado 80309

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:

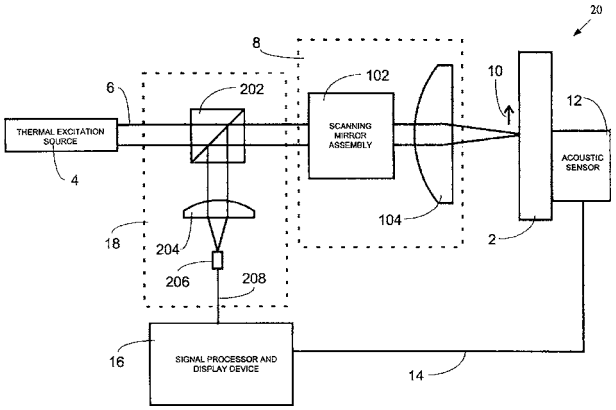
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN ERDREICH, *Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07050*
SEAN A. FULOP, *California State University, Fresno, 5245 North Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Avenue South, Eden Prairie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., P.O. Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,357,029

43.35.Sx THERMAL-ACOUSTIC SCANNING SYSTEMS AND METHODS

Robert A. Falk, assignor to Optometrix, Incorporated
15 April 2008 (Class 73/643); filed 25 March 2005

Various techniques have been devised to detect subsurface defects in materials. For example, a pulsed laser produces momentary stress and strain in a localized heated area, and sound is thereby generated. The acoustic propagation properties can then be analyzed to identify flaws. This patent



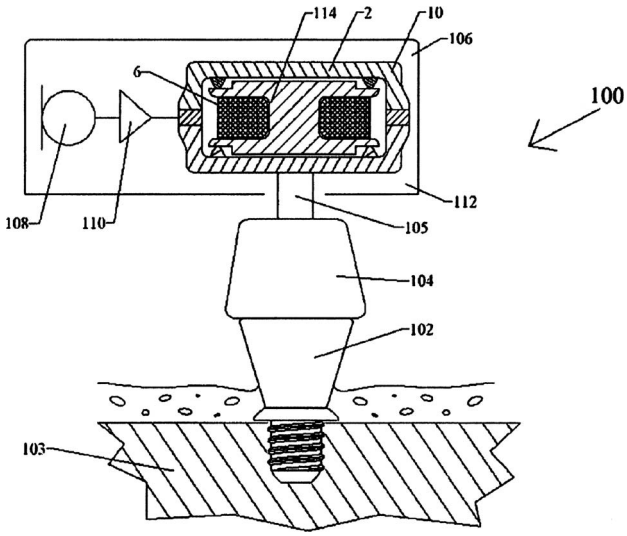
suggests that a continuous scanned laser beam can perform a similar function for testing nonuniform structures, such as electronic circuit boards with components soldered in place. One or more acoustic sensors 12 pick up pops and pings produced by the particular component under thermal stress at a given time. The signal is processed and displayed for evaluation.—GLA

7,376,237

43.38.Dv VIBRATOR FOR BONE-CONDUCTION HEARING

Patrik Westerkull, assignor to Oticon A/S
20 May 2008 (Class 381/326); filed 2 September 2004

Mechanical design details are provided for a bone conduction vibrator that is said to be more durable and easier and less costly to manufacture. The



numbers of magnets, air gaps, and coils is reduced, and a magnet is positioned so that static magnetic flux passes through first and second airgaps in parallel with the axial direction of the vibrator.—DAP

7,372,346

43.38.Fx ACOUSTIC RESONATOR

Hendrikus A. C. Tilmans and Wanling Pan, assignors to Interuniversitair Microelektronica Centrum (IMEC)
13 May 2008 (Class 333/187); filed in the European Patent Office 24 December 2003

This patent discloses some uses of air gaps in a piezoelectric element to tune the resonant frequency of a film bulk acoustic resonator. The essential concept in all cases is that one can create an effective variable series capacitance by allowing for an air gap between the piezoelectric element and the electrodes. The authors then illustrate a few ways in which these gaps may be varied by the application of a voltage, thus shifting the resonant frequency slightly. While this patent addresses an important need for a tunable rf filter, none of these discussed in the patent appear to be practical due to stability problems, vibration sensitivity, or temperature stability.—JAH

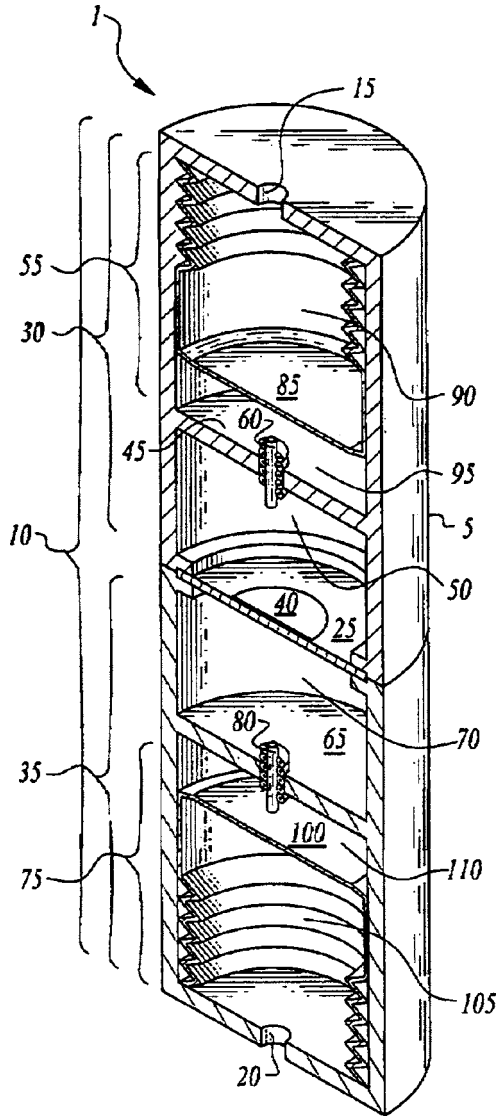
 0001-4966/2008/124(5)/2667/13/$23.00 © 2008 Acoustical Society of America

7,380,459

43.38.Fx ABSOLUTE PRESSURE SENSOR

Itzhak Sapir, assignor to Irvine Sensors Corporation
3 June 2008 (Class 73/704); filed 16 January 2007

The author discloses an absolute pressure gauge design that utilizes the principle of the gas spring to alter the resonant frequency of a membrane. In the device shown in the figure, membrane 40 symmetrically divides the gas-filled chambers 50 and 70. They are connected via pressure-equalizing leaks 60 and 80 to corresponding chambers 95 and 110, where the pressure is equalized to the outside through bellows 85 and 100. The mirror symmetry of this device is unusual, but is



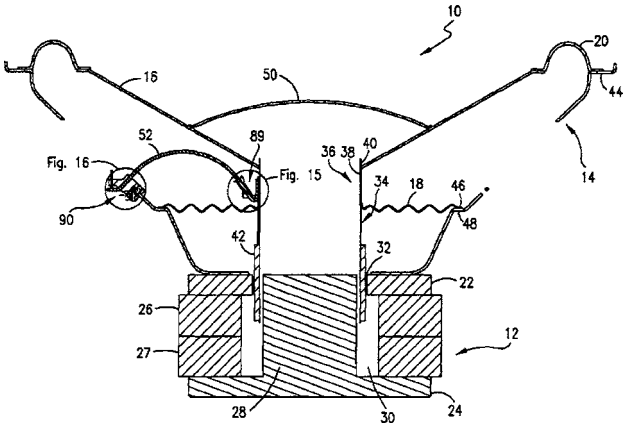
a design feature that biases the sensitivity toward absolute pressure. The device makes use of the principle of the gas spring in that the compressibility of the gas on either side of the central dividing membrane is dependent on the pressure of the gas in the cavities. By measuring the resonant frequency of a piezoelectric driver on the separating membrane 40, the pressure of the surrounding gas can be inferred.—JAH

7,356,157

43.38.Ja LOUDSPEAKER LEAD WIRE MANAGEMENT SYSTEM

Lucio Proni, assignor to JL Audio, Incorporated
8 April 2008 (Class 381/395); filed 21 December 2004

Lead wire guide bracket assemblies at 89 and 90 allow for better control of the lead wire arc at the voice coil former 36 and at the terminal on



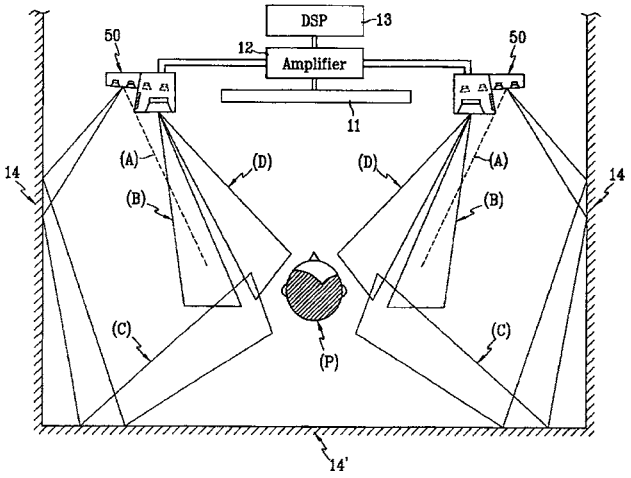
frame 14, the lead wire length, and also eases the manufacture of the loudspeaker assembly by allowing certain processes to be performed before the moving parts, including but not limited to 16, 18, 20, 36, and 52, are mounted to the frame.—NAS

7,362,874

43.38.Ja SPEAKER SYSTEM HAVING A FRONT SPEAKER INTEGRATED WITH A REFLECTION-TYPE SURROUND SPEAKER

Jin Sub Lee *et al.*, assignors to Samsung Electronics Company, Limited
22 April 2008 (Class 381/307); filed in Republic of Korea 30 December 2002

Two conventional loudspeakers reproduce left and right stereo signals. Surround speakers 50 are located adjacent to the main speakers but somewhat behind, so that their direct sound path A is shadowed by the L-shaped cabinet. Thus, surround sound C reaches the listener after being reflected by room walls 14. Considering practical dimensions for such a system, it is



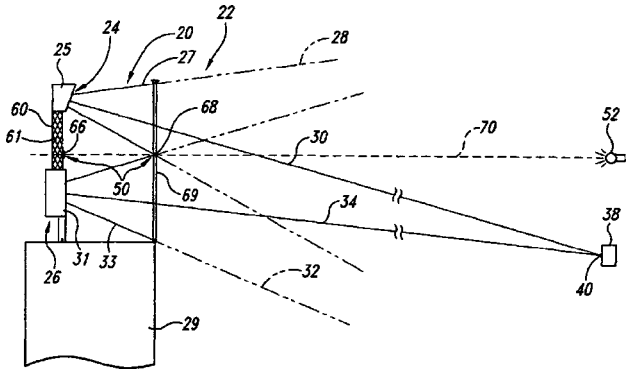
likely that the desired effect will be restricted to frequencies above 2 kHz or so. Alternatively, digital processing can be used to cancel direct sound from the surround speakers. In either embodiment the patent claims apply only to a combined loudspeaker, amplifier, and video screen, all housed in a single cabinet.—GLA

7,367,423

43.38.Ja SPEAKER ASSEMBLY WITH AIMING DEVICE

Barron Ferrell *et al.*, assignors to QSC Audio Products, Incorporated
6 May 2008 (Class 181/187); filed 25 October 2004

A high frequency horn assembly is fitted with a rudimentary gun sight to facilitate accurate aiming. According to the patent abstract, the installation technician must be able to "visually see." If that requirement is met, he



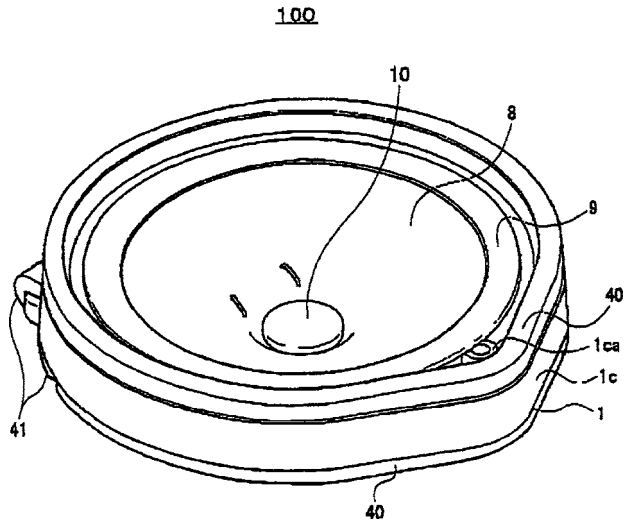
or she simply looks through the sight and adjusts the assembly until target 52 is visible. It is stated that perforated motion picture screens are sufficiently transparent for the device to be used in cinema installations.—GLA

7,377,358

43.38.Ja FRAME FOR SPEAKER DEVICE AND SPEAKER DEVICE

Hiromitsu Sasaki and Yutaka Moriyama, assignors to Pioneer Corporation
27 May 2008 (Class 181/171); filed in Japan 17 February 2005

Automotive loudspeakers must be installed quickly and easily, yet provide a good air seal. This patent describes a loudspeaker assembly that hooks into a wide clip 41 and fastens with a single screw through hole 1ca.



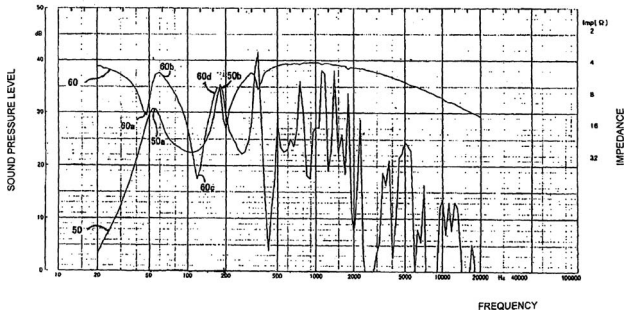
The patent text defies comprehension, but the design shown is also intended to minimize the outside frame diameter.—GLA

7,379,556

43.38.Ja SPEAKER DEVICE

Koji Maekawa and Hiroyuki Dohi, assignors to Pioneer Corporation 27 May 2008 (Class 381/338); filed in Japan 28 March 2002

This reviewer may have missed something here, but what is disclosed in the patent appears to be more of an observation than an invention. Moreover, the description is confusing, the analysis is faulty, and the results are nothing to brag about. Would you buy a subwoofer with the frequency response shown? (Points labeled 50 lie on the response curve and those labeled 60 lie on an upside-down impedance curve.) What is described is a vented box loudspeaker system having a relatively small air chamber and a



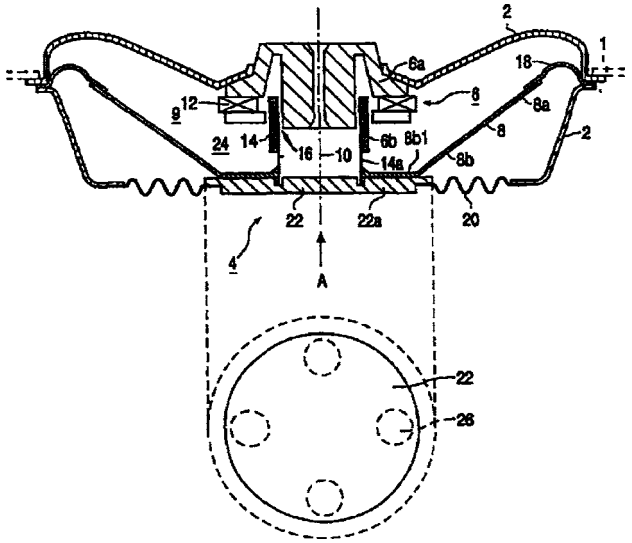
relatively long vent. In this common configuration, vent pipe resonances may degrade performance within the desired frequency range. However, when used as a subwoofer, it is fairly easy to obtain flat response from the fundamental (vent) resonance frequency up to the first pipe resonance (which, contrary to the patent text, is not the fundamental resonance of the pipe "by itself"). In a practical small subwoofer, these two frequencies may define a bandwidth of more than two octaves. Achieving the performance shown in this patent takes real effort.—GLA

7,382,893

43.38.Ja LOUDSPEAKER WITH INVERTED CONE

Jozef Arnold Frans Baeten, assignor to PSS Belgium N.V.
3 June 2008 (Class 381/416); filed in the European Patent Office 16 August 2002

The novelty of this shallow inverted cone loudspeaker lies in "bridging element" 22, which takes the form of a rigid disk. It is said to reduce overall



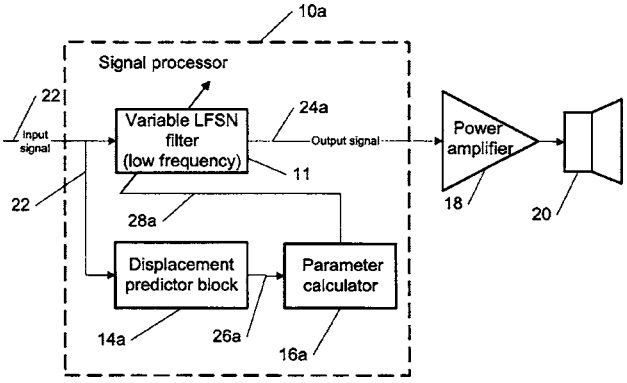
depth and can also serve as a heat radiator.—GLA

7,372,966

43.38.Lc SYSTEM FOR LIMITING LOUDSPEAKER DISPLACEMENT

Andrew Bright, assignor to Nokia Corporation
13 May 2008 (Class 381/55); filed 19 March 2004

At low frequencies, loudspeakers can suffer damage from excessive cone movement if driven too hard. This problem must be dealt with in musical instrument loudspeakers and concert sound systems, but it can also show up in small portable devices such as cellular phones. The maximum



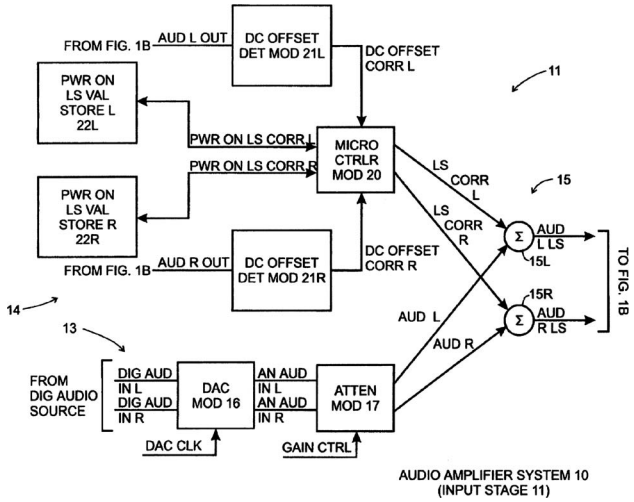
allowable voltage drive is not easy to quantify because it is a frequency-related function specific to a particular loudspeaker in a particular enclosure. The method described in this patent simplifies the process by relating the coefficients of protective shelving and notch filters to the resonance frequency and *Q* of the loudspeaker.—GLA

7,378,903

43.38.Lc SYSTEM AND METHOD FOR MINIMIZING dc OFFSET IN OUTPUTS OF AUDIO POWER AMPLIFIERS

L. Dexter Bates, assignor to Sonos, Incorporated
27 May 2008 (Class 330/9); filed 7 June 2004

Class D (switching) audio power amplifiers are prone to audible turn-on transients because the circuits required for dc offset correction require a certain amount of time to stabilize. This patent describes a fairly



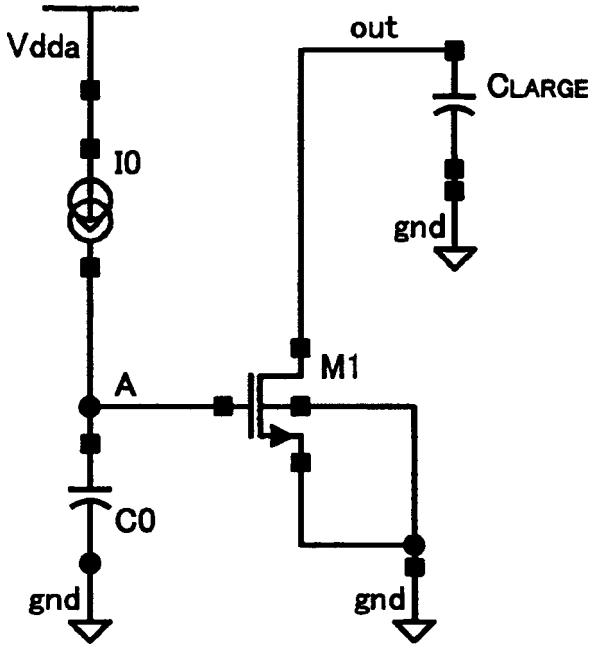
complex computerized method for modifying the input signal so as to compensate for dc offset at the time the amplifier is powered on.—GLA

7,382,187

43.38.Lc SEMICONDUCTOR INTEGRATED CIRCUIT AND AMPLIFIER FOR SUPPRESSING POP SOUND WHILE MINIMIZING VOLTAGE TRANSITION SETTLING TIME

Katsuhiko Aisu, assignor to Ricoh Company, Limited
3 June 2008 (Class 330/151); filed in Japan 27 January 2003

For analog audio amplifiers, a conventional pop reduction circuit requires a long settling time because the control voltage itself has an appreciable turn-on transient. In the circuit shown here, a small capacitor C0 is charged or discharged through constant current source I0. This creates a



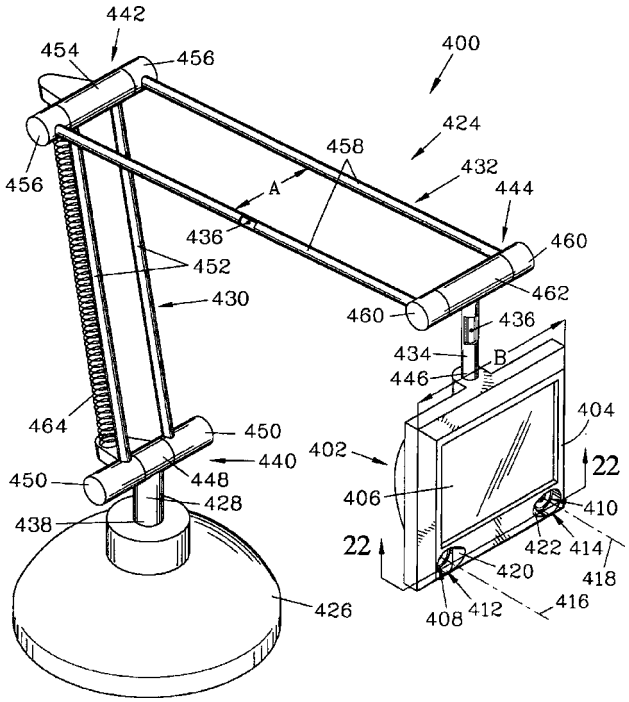
linearly changing voltage, which is applied to the gate of metal oxide semiconductor transistor M1. The smoothly changing current through Ml charges or discharges outboard capacitor Clarge to provide the required control voltage.—GLA

7,369,672

43.38.Md PERSONAL AUDIO VISUAL SYSTEM

Bruce D. Hirschhorn, Boonton, New Jersey
6 May 2008 (Class 381/333); filed 9 June 2003

Adjustable extension mounts are often used for computer monitors so that the screen can be moved and angled as desired. If the monitor includes built-in speakers then the assembly might well be described as a personal audio visual system, the title of this patent. However, two mildly novel features are included: First, the adjustable arms provide passages for concealed wiring, and second, the housing can also *contain* a light source to provide a reading light. These may not have been enough for the patent



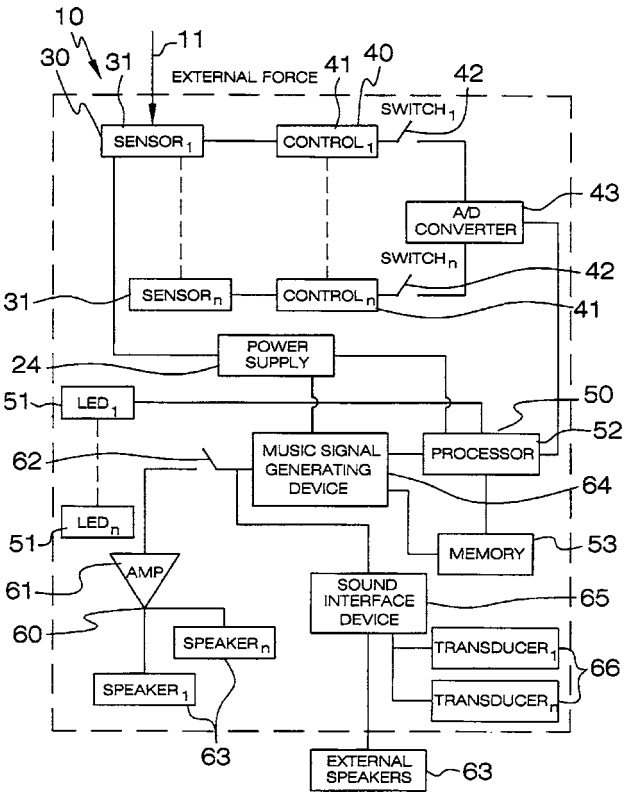
examiner because the claims are mostly devoted to detailed descriptions of every bracket, joint, and spring used in several embodiments.—GLA

7,381,884

43.38.Ne SOUND GENERATING HAND WEAR

Yourik Atakhanian, Hillsdale, New Jersey
3 June 2008 (Class 84/723); filed 3 March 2006

The Theramin was probably the first "air" instrument—one that is played without direct contact. Now imagine that each finger of a performer's hand controls its own Theramin. This patent proposes a musical glove that includes sensors at the fingertips, knuckles, and palm. The glove also houses analog-to-digital converters, a digital processor, power supply, (batteries?), an audio power amplifier, and loudspeakers. Considering the miniaturization of today's cellular phones, such an assembly is probably practical but the patent gives no hint that a working prototype has actually been constructed.—GLA

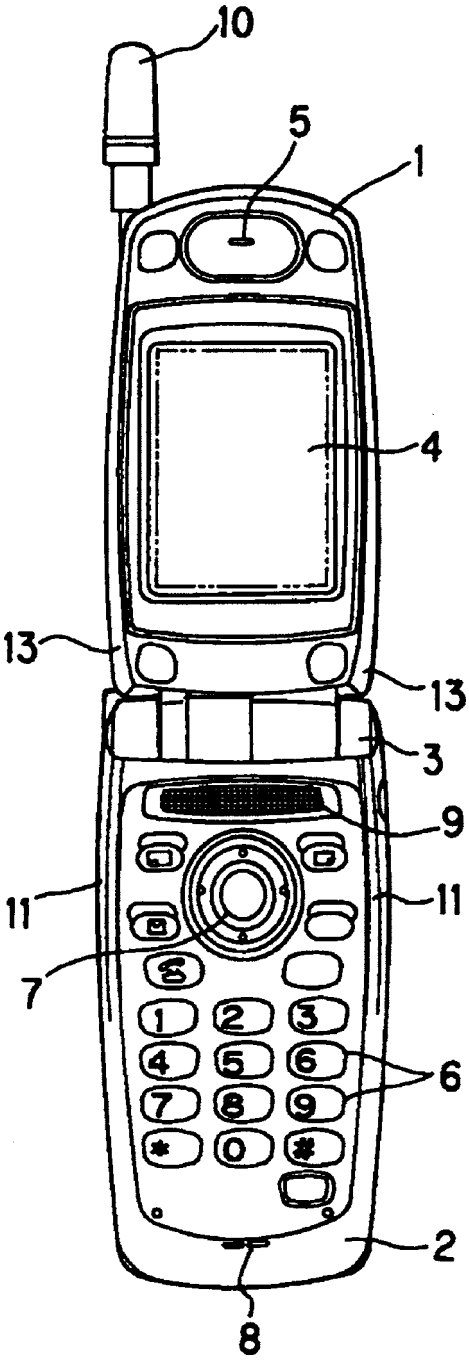


7,376,447

43.38.Si MOBILE TELEPHONE

Mamoru Yoshida *et al.*, assignors to Matsushita Electric Industrial Company, Limited
20 May 2008 (Class 455/569.1); filed in Japan 19 October 2001

A folding cellular telephone obviously must emit an audible ring signal whether the case is open or closed. Also, a user may want to receive other information such as time signals and weather forecasts without opening the



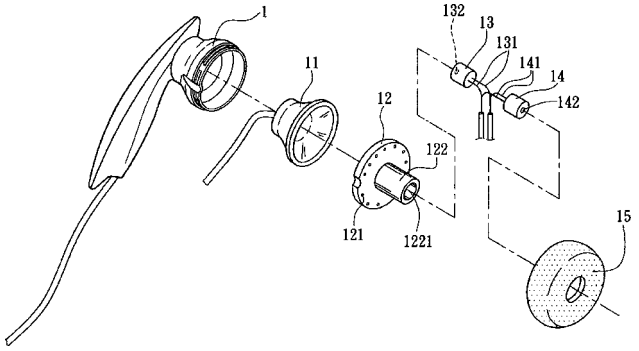
phone. Recent patents have addressed the problem by providing additional loudspeakers, sound transmission openings, acoustic valves, and the like. In the design shown, the two halves of the case are slightly separated when closed, allowing sound to emerge around the perimeter. Offset flanges hide the slot.—GLA

7,362,876

43.38.Si INTERFERENCE-FREE TRANSMITTING RECEIVING EARSET

Jin-Chou Tsai, Taipei City, Taiwan
22 April 2008 (Class 381/380); filed 27 September 2004

This mysterious in-the-ear transmitter/receiver includes loudspeaker 11 and two microphones 132 and 142, connected as shown. Sound energy in the ear canal presumably includes the user's voice as well as the sound from the loudspeaker. The microphones are mounted, "… in reversed directions to pick up sound waves from the user's external auditory canal and sound



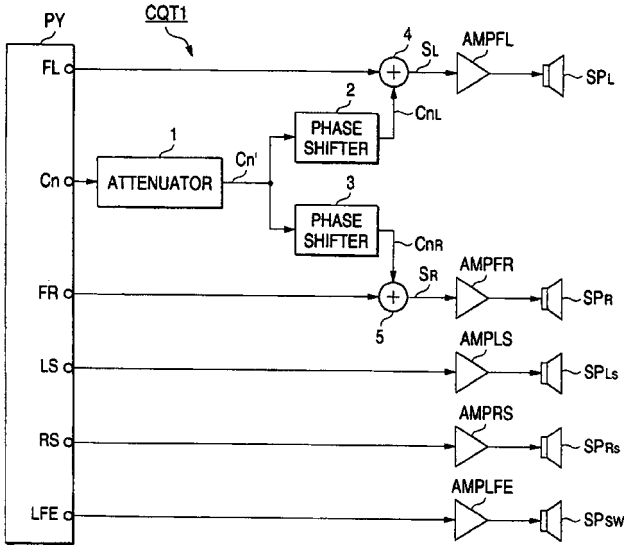
waves from the speaker, respectively, and to eliminate interference of sound waves from the the speaker with sound waves produced by the user in the external auditory canal." A second embodiment has only one microphone, plus a skirt that shields the microphone from the loudspeaker while somehow allowing sound to skip around the microphone and reach the inner ear.—GLA

7,369,666

43.38.Vk AUDIO REPRODUCING SYSTEM

Kazuhito Tatsuta *et al.*, assignors to Pioneer Corporation
6 May 2008 (Class 381/302); filed in Japan 10 August 2001

When reproducing 5.1 surround sound (or, for that matter, a two-channel stereo) most automotive sound systems rely on the left and right speakers to produce a phantom center image. Numerous circuits have been



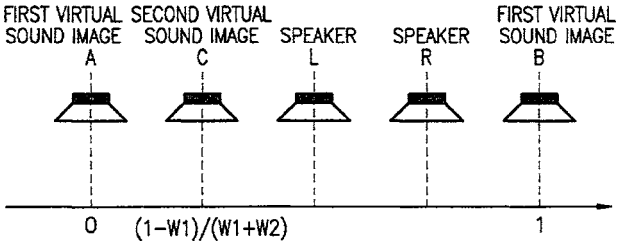
designed to adjust the relative delay and intensity of the two speakers so that an accurate sound image can be heard by the driver. Providing satisfactory sound images for both the driver and a front passenger is trickier. This patent asserts that in both locations the binaural correlation coefficient is negative in the range from about 200 to 400 Hz. Therefore, if this range is phase-shifted at 180°, both listeners will enjoy an improved center image. Several embodiments are described in the patent.—GLA

7,382,885

43.38.Vk MULTICHANNEL AUDIO REPRODUCTION APPARATUS AND METHOD FOR LOUDSPEAKER SOUND REPRODUCTION USING POSITION ADJUSTABLE VIRTUAL SOUND IMAGES

Sang-wook Kim *et al.*, assignors to Samsung Electronics Company, Limited
3 June 2008 (Class 381/17); filed in Republic of Korea 10 June 1999

A lot of engineering effort has been devoted to the problem of generating satisfactory five-channel surround sound from a pair of loudspeakers. Most schemes include some combination of fixed head-related transfer functions (HRTFs) and interaural crosstalk cancellation. This patent argues that there is a need for adjustable virtual image locations. This feature (which has been extensively studied in the field of virtual reality audio reproduction) usually requires storing several sets of HRTFs and then interpolating



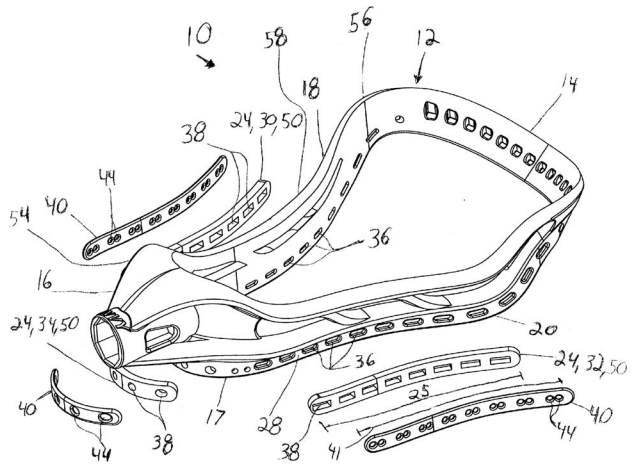
between pairs to generate intermediate locations. The patent asserts that, in the front hemisphere at least, there is an easier way. It is well known that a convincing virtual image can be located at any point between the front left and right loudspeakers by suitably delaying and scaling the left and right versions of a mono input signal. Therefore, it should be possible to generate a new virtual sound image between two virtual sound images by using the same technique.—GLA

7,344,460

43.40.Kd LACROSSE HEAD WITH CUSHIONED SIDEWALLS

Paul Gait, assignor to Rawlings Sporting Goods Company, Incorporated
18 March 2008 (Class 473/513); filed 1 November 2006

Energy absorbing elements 24 are placed between securing elements 40 and edgewalls 18 and 20 of scoop 14. Prior art attributable to the



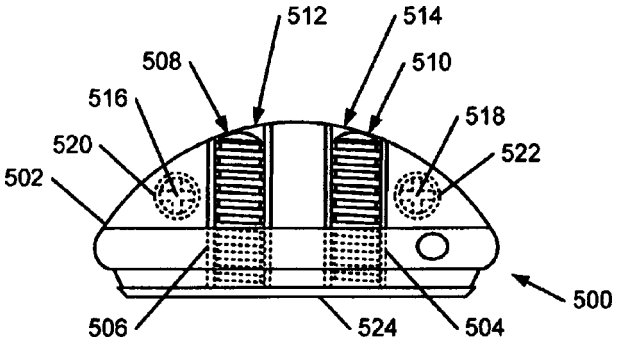
Algonquin, Iroquois, and other early practitioners is not cited although the patent states that parts of the patent are protected under US copyright law.—NAS

7,354,355

43.40.Sk GOLF CLUB HEAD OR OTHER BALL STRIKING DEVICE WITH MODIFIABLE FEEL CHARACTERISTICS

Gary G. Tavares *et al.*, assignors to Nike, Incorporated
8 April 2008 (Class 473/329); filed 1 October 2004.

The "give" and "feel" of putter head 500 can be adjusted by the retailer, pro, or duffer by changing weighted members 508, 510, 516, and 518, which are fitted into receptacles 512, 514, 520, and 522. The sound made when the club face 524 strikes the ball may be altered by the choice of, and



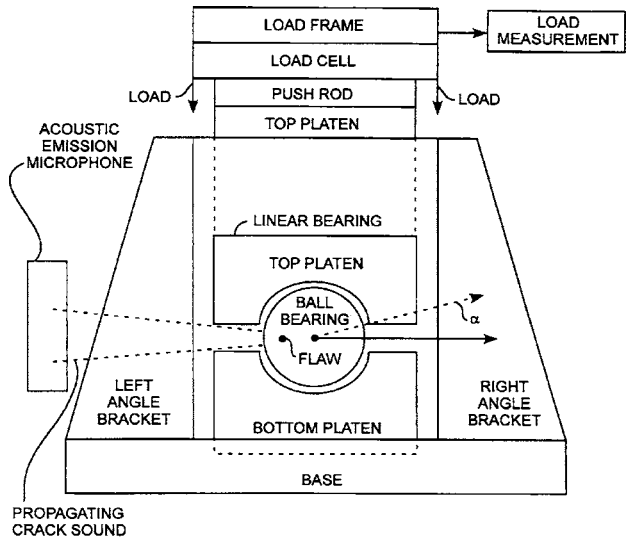
tightness of fit, of feel members 508 and 510. This reviewer wishes his game was to the point where tuning of the striking device 500 was of greater concern than his score.—NAS

7,370,537

43.40.Le CERAMIC BALL BEARING ACOUSTIC TEST METHOD

Michael J. O'Brien *et al.*, assignors to The Aerospace Corporation
13 May 2008 (Class 73/818); filed 15 November 2004

It is difficult to detect minute cracks in brittle ball bearings. This patent includes a good summary of prior art, including ultrasound, x ray, and fluorescent dye testing. A new method is described in which a bearing is compressed in a test jig, producing tensile stresses that promote the growth of



FRACTURE SOUND PROOF TEST FIXTURE

cracks and flaws. This process is accompanied by characteristic sounds that can be detected and analyzed.—GLA

7,364,022

43.40.Tm CONTROLLABLE MAGNETORHEOLOGICAL FLUID DEVICES FOR MOTION DAMPING

Faramarz Gordanihejad *et al.*, assignors to University of Nevada
29 April 2008 (Class 188/267.2); filed 4 April 2005

The damper described in this patent consists of a piston, which divides the volume within a cylinder into two portions. The two volumes are filled with a magnetorheological fluid and communicate with each other via orifices and passages within the piston or outside of the cylinder. Electromagnetic coils are provided that can change the magnetic fields in the flow passages, thus changing the effective viscosity of the magnetorheological fluid there and producing changes in the damping characteristics.—EEU

7,364,221

43.40.Tm REDUCTION OF VIBRATION TRANSFER

Laurent Tahri *et al.*, assignors to Henkel Kommanditgesellschaft auf Aktien
29 April 2008 (Class 296/187.02); filed in the European Patent Office 6 October 2005

Propagation of vibration and structure-borne sound along hollow structural elements is impeded by adding damping treatments to the interior surfaces of these elements. This patent describes materials and a novel way for inserting these treatments into the structural elements. The damping materials are applied around forms that fit loosely into the structural elements. The forms with the damping materials are inserted into these elements, and then heat is applied to expand the damping materials so that they make good contact with the inside surfaces of the structural elements.—EEU

7,367,435

43.40.Tm SHOCK ABSORBER DESIGN METHOD

Hiroshi Yamauchi *et al.*, assignors to Mitsubishi Jidosha Kogyo Kabushiki Kaisha
6 May 2008 (Class 188/322.22); filed in Japan 31 May 2002

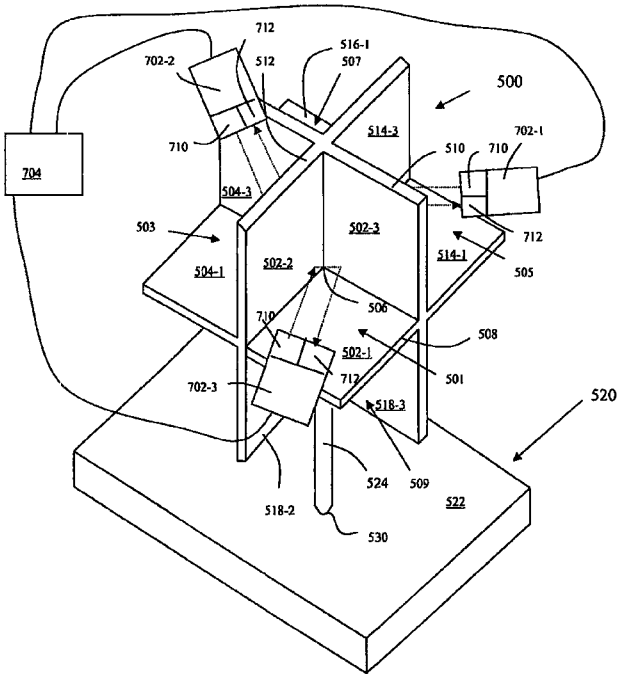
In piston-type shock absorbers for automotive applications there may occur self-excited vibrations due to slight cocking of the piston, locking it briefly against the cylinder. This repeated sticking results in a stick-slip motion and an attendant noise. This patent in essence presents relations between the friction coefficient, piston radius, and piston thickness that lead to avoidance of the self-excited vibrations due to cocking and also describes favorable geometries for the cylindrical surface of the piston.—EEU

7,374,296

43.40.Tm REFLECTIVE CORNER CUBE ARRAY

George D. Hadden, assignor to Honeywell International Incorporated
20 May 2008 (Class 359/529); filed 12 August 2005

An array 508 consisting of three mutually orthogonal reflective surfaces, such as 502-1, 502-2, and 502-3, is attached at a measurement point 530 on a vibrating object 520. Laser Doppler velocimeters, such as 702-3, each containing a laser beam generating circuit 710 and a detection circuit



712, are used to measure the motion of the reflective array in the directions of the laser beams. Use of three noncoplanar laser arrangements can provide information on the test object's vibrations along three axes, regardless of the orientation of the reflective array—EEU

7,377.339

43.40.Tm SYSTEM AND METHOD FOR DAMPING VIBRATION IN A DRILL STRING

Mark Ellsworth Wassell *et al.*, assignors to APS Technology, Incorporated
27 May 2008 (Class 175/322); filed 19 April 2007

Those interested in the applications of magnetorheological (MR) fluids to mechanical damping may find this patent interesting. The authors disclose an idea for a damper section of a drill string using the said fluid to damp the vibrations of a drill bit. The damper comprises an upper and a lower section, the upper containing a MR fluid in a cylindrical layer between an inner driving shaft and an outer driven shaft. They design the walls between these cylinders with helical grooves that trap ball bearings in the gap, allowing torque to be transmitted to a lower section of the damper while enabling axial sliding motions. The lower or down-hole section of the damper consists of a variable-spring-rate stack of Belleville washers, so that the resonant frequency of the damper can be tuned by applying more or less axial force. The damping is controlled by electromagnets that get control signals from the surface in response to linear position sensors on the damper. The whole assembly is quite complex; about 30 diameters long (probably 20 ft by today's drilling standards) and containing hundreds of specially machined parts.—JAH

7,380,804

43.40.Tm VIBRATION DAMPING APPARATUS FOR STEERING SYSTEMS

Jong-Gil Lee, assignor to Kia Motors Corporation
3 June 2008 (Class 280/90); filed in Republic of Korea 17 October 2005

The object of the apparatus described in this patent is reduction of the vibrations that are transmitted from a vehicle's wheels to its steering wheel. A lower shaft is coupled to the gear box and an upper shaft is connected to the steering wheel. Each of these shafts is splined to a set of blades that are located in a fluid-filled chamber and that do not contact each other directly. One set of these blades has orifices whose openings are controlled actively, based on the rotational displacement of the steering wheel, permitting the steering wheel to actuate the steering mechanism while it is isolated from wheel vibrations.—EEU

7,382,327

43.40.Tm ANTENNA VIBRATION ISOLATION MOUNTING SYSTEM

Larry A. Nelson, assignor to The Boeing Company
3 June 2008 (Class 343/709); filed 17 November 2005

Maintenance of an antenna's pointing accuracy requires minimizing its rotational vibrations. The mounting system described in this patent is intended to accomplish this by making use of three spring-supported stages, each provided with guides that permit only translational motion. The first stage permits only vertical translation relative to the supporting structure, the second permits only horizontal motion relative to the first in a given direction, and the third—on which the antenna is mounted—permits only horizontal motion relative to the second in the direction perpendicular to that of the second.—EEU

7,382,510

43.40.Tm ACTUATOR

Mitsuhiro Yoda, assignor to Seiko Epson Corporation
3 June 2008 (Class 359/224); filed in Japan 5 September 2003

The purpose of the actuator described in this patent is to oscillate a mirror in an optical scanner at a high frequency. This miniature device makes clever use of a two-stage dynamic arrangement. The mirror constitutes one rotational inertial element and is supported from a second inertial element via a flexible shaft. The second element is supported from a base via a second flexible shaft. The inertias and shaft stiffnesses are chosen so that at the driving frequency the mirror oscillates with a large amplitude while the second inertia element moves very little.—EEU

7,372,265

43.40.Vn COMPENSATION OF MAGNETIC FIELD DISTURBANCES DUE TO VIBRATIONS IN A MAGNETIC RESONANCE IMAGING SYSTEM

Cornelis Leonardus Gerardus Ham *et al.*, assignors to Koninklijke Philips Electronics N.V.
13 May 2008 (Class 324/307); filed in the European Patent Office 5 February 2003

Mechanical vibrations that can degrade the performance of a magnetic resonance imaging (MRI) system are caused primarily by (1) the varying Lorentz forces that act on the gradient magnetic coil system (operation of which controls the scanning process) due to interaction of the varying current in this system with the main magnetic field, and by (2) the Lorentz forces that are exerted on other parts of the MRI system as the result of interaction of the main magnetic field with eddy currents induced in these parts by the varying magnetic field. The present patent describes a feed-forward control system in which compensation of the magnetic field is determined on the basis of the actual currents in the gradient coil system.—EEU

7,373,262

43.40.Vn METHOD FOR IMPLEMENTING VIBRATION SUPPRESSION SYSTEM UNDER CONTROL FROM A REMOTE SITE

Baruch Pletner and Gregory A. Zvonar, assignors to IPTrade, Incorporated
13 May 2008 (Class 702/56); filed 27 October 2005

Because precision equipment cannot be designed a priori to meet all of its vibration requirements, vibration sensors and actuators are installed at selected locations in such equipment and active control using various algorithms is employed to limit the undesirable vibrations. In order to avoid the practical difficulties associated with hard-wired connections of the vibrating parts to the control circuits, especially in rapidly moving or rotating items, a system is provided that transmit sensed data to a remote site, where a controller is located, and control signals are transmitted from that site to the actuators. This patent describes the functional blocks of appropriate systems.—EEU

7,357,031

43.40.Yq PRECISION SENSOR ASSEMBLY FOR ROTATING MEMBERS IN A MACHINE TOOL

Gianni Trionfetti and Andrea Guidotti, assignors to Balance Systems S.r.l.
15 April 2008 (Class 73/649); filed in the European Patent Office 8 March 2004.

Axial vibrations of the rotor in a high-speed precision grinder that are indicative of contact of the grinding wheel with a workpiece are detected by means of a sensor located on the rotor and disposed symmetrically about the rotor's axis. The sensor consists of a piezoelectric element to which a mass is attached via an elastic member, with the mass and elastic member selected to amplify vibrations in a desired frequency range. The signal generated by the piezoelectric element is fed to a coil, which also is attached to the rotor and is symmetric about its axis, and which is disposed adjacent to a similar coil that is mounted on the stator. The signal received by the latter is amplified and fed to an external regulator.—EEU

7,365,530

43.40.Yq METHOD AND APPARATUS FOR VIBRATION DETECTION

James M. Bailey *et al.*, assignors to Allegro Microsystems, Incorporated
29 April 2008 (Class 324/207.25); filed 16 September 2004

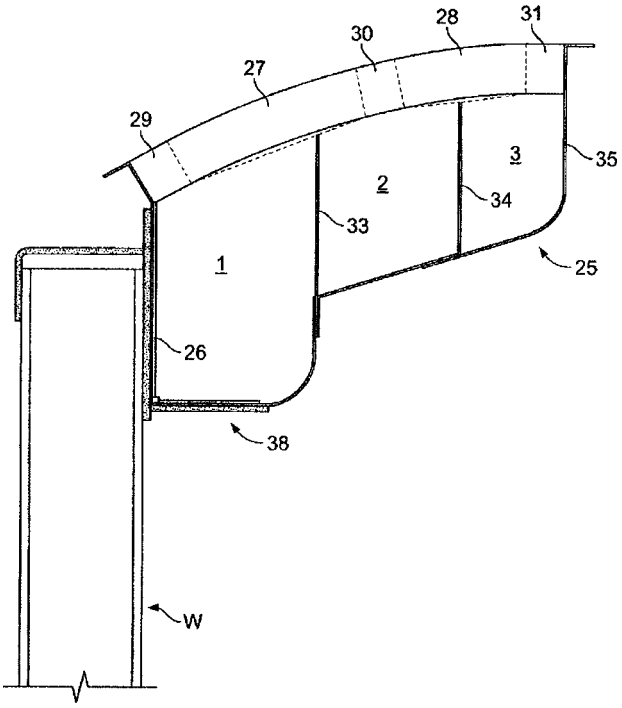
This apparatus, intended for detecting lateral and rotational vibrations in antilock brake systems and the like, makes use of a plurality of magnetic field sensors and two or more rotation detectors. The resulting signals may be processed in various ways and fed to a controller in combination with signals from detectors for speed and direction, air gap, and temperature.—EEU

7,380,636

43.50.Gf NOISE REDUCING EQUIPMENT

Hiroshi Yano *et al.*
3 June 2008 (Class 181/210); filed 10 August 2005

A series of Helmholtz resonators is attached to an extension that sits on top of a sound barrier wall. These chambers upgrade the benefit of the wall



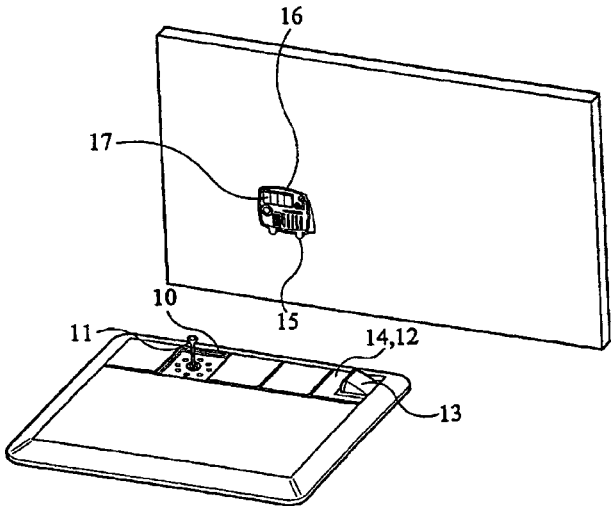
at specified low frequencies.—CJR

7,344,446

43.60.Bf BALL DETECTION APPARATUS FOR DETECTING WHEN A GOLF BALL IS HIT OFF A GOLF TEE

Martin Wyeth, Wiltshire, United Kingdom
18 March 2008 (Class 473/192); filed in United Kingdom 28 March 2003

The intensity of the bandpass filtered signal from directional microphone 15, aimed at golf tee 11, is compared to the club speed measured by radar gun 13, which is aimed slightly above the top of the tee. When the maxima from the two signals are within a certain time window, the device determines that ball has been ″hit off″ the tee on automatic gold ball teeing machine 10. Maximum club speed can be shown on display 16, which can



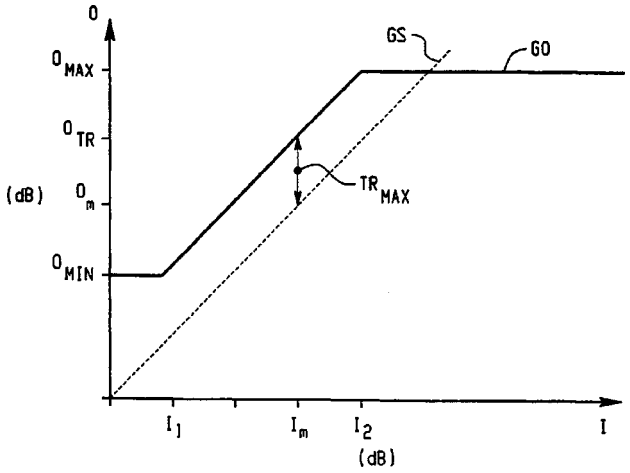
be mounted to control panel 17, which may also house the signal processing circuitry.—NAS

7,372,969

43.66.Ts METHOD FOR PROCESSING AN INPUT SIGNAL TO GENERATE AN OUTPUT SIGNAL, AND APPLICATION OF SAID METHOD IN HEARING AIDS AND LISTENING DEVICES

Hans-Ueli Roeck, assignor to Phonak AG
13 May 2008 (Class 381/317); filed 16 September 2005

This patent describes a method of limiting the output of a hearing aid to prevent wearer annoyance from transient sounds such as dishes and cutlery clattering. The method includes determining an adaptive output limiting



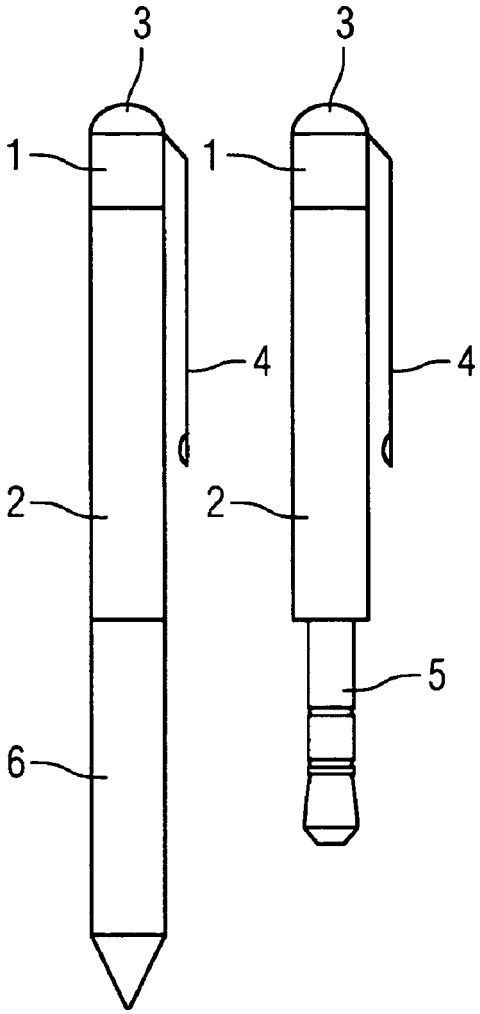
threshold as a function of the running mean value of the input signal and the compression ratio.—DAP

7,372,970

43.66.Ts POCKET HEARING AID

Gerhard Ach-Kowalewski, assignor to Siemens Audiologische Technik GmbH
13 May 2008 (Class 381/322); filed in Germany 26 February 2004

The wireless version of a pocket hearing aid. The normally used cable between the body and earpiece portions of the pocket hearing aid is eliminated. The body portion resembles a pen, and the earpiece portion is an in-the-ear or behind-the-ear hearing aid with a wireless communication link,



which could have a detachable wireless receiver. The microphone assembly has on one end an antenna shaped as a pocket clip and on the other end a jack to plug into audio input connectors in classroom, stereo, and mobile applications, thus eliminating the need for an accessory shoe.—DAP

7,372,971

43.66.Ts HEARING DEVICE WITH FUEL CELL

Hans-Ueli Roeck, assignor to Phonak AG
13 May 2008 (Class 381/323); filed 12 January 2004

A fuel cell, which may be a hydrogen-oxygen type, provides power for at least the digital signal processor in a hearing device. The fuel tank to store the fuel fluid may be made of a flexible material and be detachable from or be contained within the hearing device.—DAP

7,372,972

43.66.Ts EAR INSERT FOR HEARING AIDS

Gerhard Ach-Kowalewski *et al.*, assignors to Siemens Audiologische Technik GmbH
13 May 2008 (Class 381/328); filed in Germany 26 February 2004

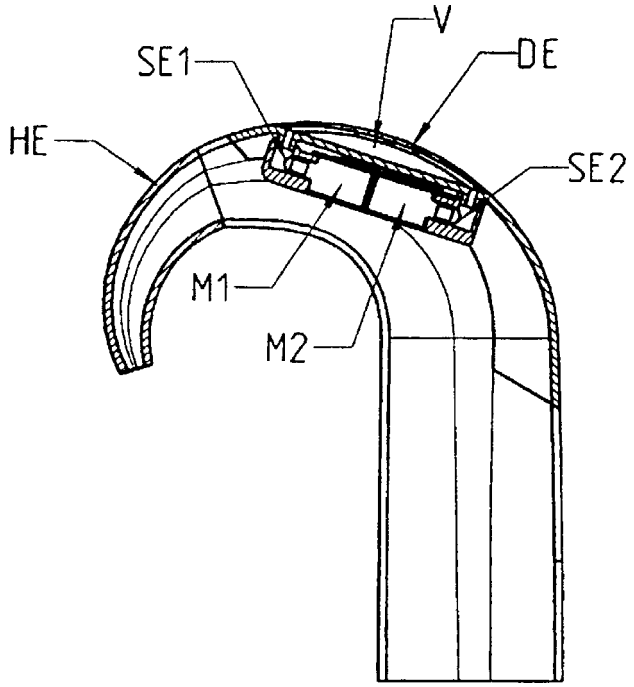
To create a miniature megaphone and improve high frequency output, an exchangeable capsule with a funnel-shaped or exponential horn is inserted into the sound outlet of an in-the-ear hearing aid or earmold, harkening back to Mead Killion's tuned earmolds of the late 1970s.—DAP

7,372,973

43.66.Ts HEARING AID

Hilmar Meier, assignor to Phonak AG
13 May 2008 (Class 381/330); filed 28 February 2003

To reduce spurious acoustic signals, also known as wind noise, that are produced by air hitting the edges of a hearing aid housing, an open-pore rigid cover of at least 0.5 mm thick is integrated into the housing over the



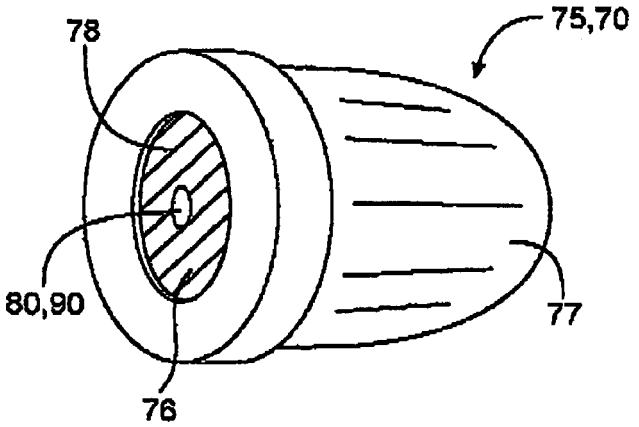
acoustic input apertures of at least two microphones. The cover is designed so that its surface is flush to the housing surface.—DAP

7,379,555

43.66.Ts PRECISION MICROHOLE FOR EXTENDED LIFE BATTERIES

Richard Gable *et al.*, assignors to Insound Medical, Incorporated
27 May 2008 (Class 381/322); filed 26 January 2005

The operating life of zinc air hearing aid batteries is typically shorter in very dry or very damp climates versus locations with moderate levels of humidity. To improve battery life for a deep-fitting hearing aid, the flow of oxygen and moisture into and out of a battery is controlled with a



diffusivity-limiting membrane. The polymer diffusion element has compressed and noncompressed regions, with the noncompressed regions aligning with the holes in the shell of the hearing aid.—DAP

7,370,655

43.66.Vt METHOD OF FORMING AN EARPLUG BY LASER ABLATION AND AN EARPLUG FORMED THEREBY

Duncan Taylor, assignor to Cabot Safety Intermediate Corporation
13 May 2008 (Class 128/864); filed 17 January 2003

This describes using laser ablation to form a deep hole along the longitudinal axis of an earplug for attachment of a cord or for forming a pattern on the surface of the plug.—JE

7,177,803

43.72.−p METHOD AND APPARATUS FOR ENHANCING LOUDNESS OF AN AUDIO SIGNAL

Marc A. Boillot *et al.*, assignors to Motorola, Incorporated
13 February 2007 (Class 704/209); filed 22 October 2002

A method for enhancing apparent loudness is presented, which exploits the processing properties of the cochlea. To wit, a sound that excites more than one cochlear spectral critical band is perceived to be louder than a sound that excites only one critical band. Therefore, it is reasoned that a speech sound whose formant bandwidths are expanded to excite more critical bands will sound louder. The speech is first coded using linear prediction, and then a filter is employed, which expands formant bandwidths by scaling the linear prediction coefficients (assumed to correspond to the formants during vowel-like sounds) using a power series. This technique has been used in the past to correct vocoder digitization errors, so this patent can only lay claim to the particular application of formant expansion.—SAF

7,181,390

43.72.−p NOISE REDUCTION USING CORRECTION VECTORS BASED ON DYNAMIC ASPECTS OF SPEECH AND NOISE NORMALIZATION

James G. Droppo *et al.*, assignors to Microsoft Corporation
20 February 2007 (Class 704/226); filed 26 July 2005

The sort of noise reduction described here requires training, and so the envisioned applications center around speech recognition. A frame of a noisy signal is converted to a "noisy feature vector" of the typical sort, such as cepstral coefficients. This noisy signal appears to be obtained from an otherwise identical clean signal, and then the pair of sets of vector mixture components obtained from these can be used in a "noise reduction trainer" to determine a "correction vector" for each Gaussian mixture component in the noisy channel speech. As usual from Microsoft, the patent is detailed, sensible, and bulky with legalese delimiting the broadest possible coverage.—SAF

7,177,805

43.72.Dv SIMPLIFIED NOISE SUPPRESSION CIRCUIT

Stephen S. Oh and Ethan T. Davis, assignors to Texas Instruments Incorporated
13 February 2007 (Class 704/225); filed 14 January 2000

Here we have another spectral noise suppression patent, which uses the now-standard general architecture of computing the fast Fourier transform of a brief windowed signal frame, computing a power estimate from some of these, smoothing the said power estimate, and computing a noise estimate, using the previous pair of functions to compute a gain function, multiplying the transformed windowed signal by the gain function, and then inverting the transform. Specific improvements in computational complexity

are claimed here for various of these steps, including computing the power as the absolute value instead of the square, and employing a very simplistic means of estimating and updating the noise spectrum. The resulting noise suppression is claimed (though not demonstrated) to be as good as traditional spectral subtraction.—SAF

7,181,393

43.72.Fx METHOD OF REAL-TIME SPEAKER CHANGE POINT DETECTION, SPEAKER TRACKING, AND SPEAKER MODEL CONSTRUCTION

Lie Lu and Hong-Jiang Zhang, assignors to Microsoft Corporation
20 February 2007 (Class 704/243); filed 29 November 2002

An audio signal is segmented into frames, from each of which a feature extractor computes a pitch feature, an eight-dimensional mel-frequency cepstral vector, and a ten-dimensional linear spectral pairs vector. In a "pre-segmentation" process, the covariance of each successive (combined) feature vector is computed, and a distance is determined from the covariance of the current segment and a previous segment. This distance is used to make an initial decision whether there is a potential speaker change between the two frames. If there may be a speaker change, a refinement process is employed in which the probability of a speaker change is computed.—SAF

7,184,958

43.72.Ja SPEECH SYNTHESIS METHOD

Takehiko Kagoshima and Masami Akamine, assignors to Kabushiki Kaisha Toshiba
27 February 2007 (Class 704/260); filed in Japan 4 December 1995

In speech synthesis, stored synthesis units must be concatenated to yield the desired output pronunciation. As a practical matter, one must have both a large inventory of stored units and also some means of modifying the stored units at least in the parameters of pitch and duration. With this much established from the prior art, this patent painstakingly takes us through a number of unintelligible specific methods for clustering the inventory of units and for modifying the units during the synthesis. After 46 figures and 16 pages of text, it is surprising to find only seven claims at the end, none of which appears to state any novel process. Perhaps the examiners got as tired as this reviewer.—SAF

7,181,392

43.72.Ne DETERMINING SPEECH RECOGNITION ACCURACY

Shailesh B. Gandhi *et al.*, assignors to International Business Machines Corporation
20 February 2007 (Class 704/235); filed 16 July 2002

"The invention disclosed herein provides a solution for determining the accuracy of a speech recognition system." Who knew that this was a major problem in the speech recognition field? The patent, by carefully invoking a variety of software bells and whistles such as annotated log files and ancillary speech recognizers, effectively lays claim to a process of comparing the text output of a speech recognizer with the audio recording that led to the said output. This is indeed the general means for determining the accuracy of a speech recognizer. It is always gratifying to learn that what one thought was so obvious, in fact is.—SAF

7,177,808

43.72.Pf METHOD FOR IMPROVING SPEAKER IDENTIFICATION BY DETERMINING USABLE SPEECH

Robert E. Yantorno *et al.*, assignors to The United States of America as represented by the Secretary of the Air Force
13 February 2007 (Class 704/246); filed 18 August 2004

A simple means of improving standard speaker identification (ID) methods in multiple-speaker or degraded speech environments is described. In order to get a reliable result from a speaker ID algorithm, it should be utilized only on the speech of one speaker at a time. This patent suggests that a speaker ID algorithm can, in fact, be used as a prescreening process, to first identify signal segments that are usable for traditional speaker ID (i.e., where one person's speech is present in a reasonable signal quality). First, speech vectors are computed from signal frames in any of the usual ways, such as cepstral coefficients. "With the knowledge of the frame distances from the speaker models, a frame of speech can be defined as usable in different ways."—SAF

7,371,218

43.80.Vj IMMERSIVE PORTABLE ULTRASOUND SYSTEM AND METHOD

Andrew L. Walston *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
13 May 2008 (Class 600/437); filed 14 January 2003

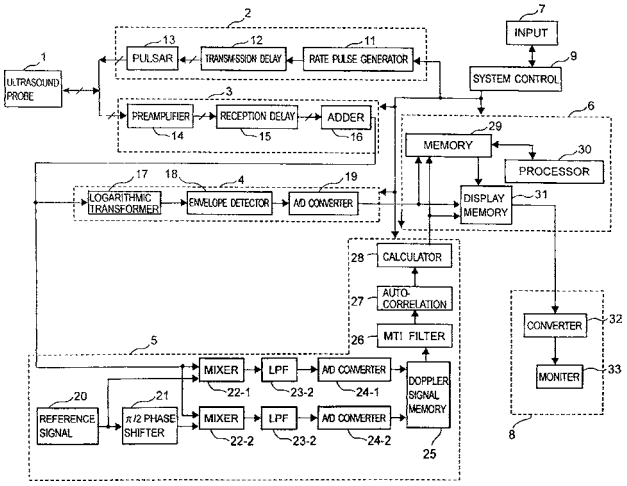
The display in this small lightweight system is adapted for viewing ultrasound images close to the eye of the user by using a magnifying lens for viewing and including a shade or eye piece to facilitate use in different lighting conditions.—RCW

7,371,219

43.80.Vj ULTRASOUND DIAGNOSIS APPARATUS OPERABLE IN DOPPLER MODE

Fumiyasu Sakaguchi and Kazuhito Nakata, assignors to Kabushiki Kaisha Toshiba
13 May 2008 (Class 600/455); filed 6 May 2004

Two Doppler-mode frames are produced from groups of ultrasound beam directions. The last group of the ultrasound beam directions in one Doppler-mode frame and the first group of ultrasound beam directions in



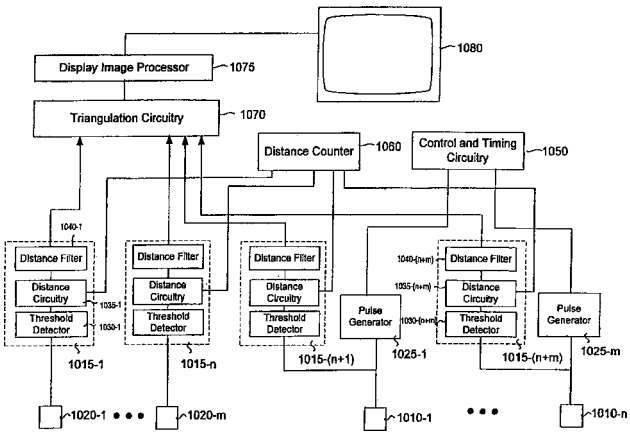
the other Doppler-mode frame are treated as one Doppler group by a controller. The Doppler signals are detected and used to form Doppler-mode images.—RCW

7,374,537

43.80.Vj PERFORMING ULTRASOUND RANGING IN THE PRESENCE OF ULTRASOUND INTERFERENCE

N. Parker Willis, assignor to Boston Scientific Scimed, Incorporated
20 May 2008 (Class 600/438); filed 27 September 2004

Distance is measured using an ultrasound pulse time of flight between two transducers. The system may contain multiple transducers from which pulses are transmitted and also corresponding multiple transducers that



receive the pulses. Because ultrasound interference in the apparatus typically yields the shortest distance measurement, interference is suppressed by taking the largest distance that is measured.—RCW

7,374,538

43.80.Vj METHODS, SYSTEMS, AND COMPUTER PROGRAM PRODUCTS FOR ULTRASOUND MEASUREMENTS USING RECEIVE MODE PARALLEL PROCESSING

Kathryn R. Nightingale and Gregg E. Trahey, assignors to Duke University
20 May 2008 (Class 600/443); filed 6 October 2003

A pulse wave emitted from an array of ultrasound transducers displaces a target region. A tracking pulse wave emitted from the transducer array is reflected from the target region. Tracking signals produced in response to the tracking pulse wave are received from locations in the target region.—RCW

7,374,744

43.80.Vj HARMONIC ULTRASOUND IMAGING WITH MICROBUBBLES

Ernest G. Schutt, assignor to Imcor Pharmaceutical Company
20 May 2008 (Class 424/9.5); filed 9 October 2002

An ultrasound pulse wave incident on gas-filled microbubbles is reradiated at frequencies other than the exciting frequency, and the reradiated energy at these other frequencies is used to produce an image.—ROW

7,374,745

43.80.Vj ULTRASOUND CONTRAST AGENTS

Gary L. Cantrell, assignor to Mallinckrodt, Incorporated
20 May 2008 (Class 424/9.52); filed 11 March 2004

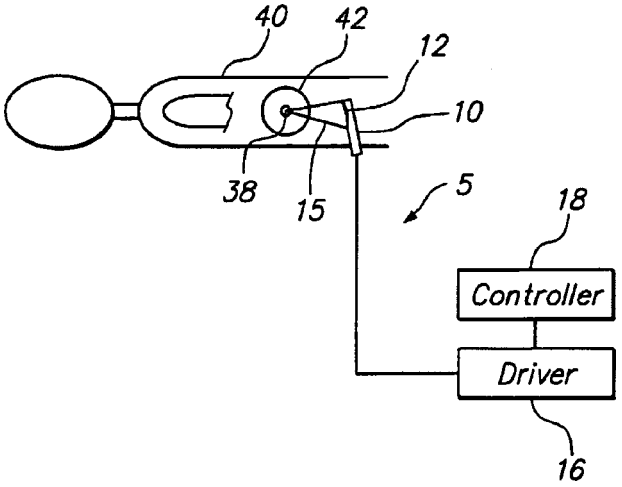
A shell that encapsulates gas microbubbles comprising these agents is made from a blend of bipolar compounds having mixed carbon chain links. The compounds have hydrophobic groups with carbon chain links greater than or equal to 16 and less than or equal to 32. The polar groups are connected to the hydrophobic groups by a linker.—RCW

7,377,900

43.80.Vj ENDOCAVITY FOCUSED ULTRASOUND TRANSDUCER

Shuki Vitek *et al.*, assignors to Insightec-Image Guided Treatment Limited
27 May 2008 (Class 600/439); filed 2 June 2003

Acoustic energy from this transducer is delivered to treat tissue at a target site. The apparatus containing the transducer has two elongated members. One member has one or more transducer elements on its distal end.



The two members are introduced sequentially and the member with the transducer elements is placed adjacent to the target site.—RCW

Methods for reconstructing acoustic quantities based on acoustic pressure measurements

Sean F. Wu
Department of Mechanical Engineering, Wayne State University, Detroit, Michigan 48202

(Received 3 December 2006; revised 25 July 2008; accepted 29 July 2008)

This paper presents an overview of the acoustic imaging methods developed over the past three decades that enable one to reconstruct all acoustic quantities based on the acoustic pressure measurements taken around a target source at close distances. One such method that has received the most attention is known as near-field acoustical holography (NAH). The original NAH relies on Fourier transforms that are suitable for a surface containing a level of constant coordinate in a source-free region. Other methods are developed to reconstruct the acoustic quantities in three-dimensional space and on an arbitrary three-dimensional source surface. Note that there is a fine difference between Fourier transform based NAH and other methods that is largely overlooked. The former can offer a wave number spectrum, thus enabling visualization of various structural waves of different wavelengths that travel on the surface of a structure; the latter cannot provide such information, which is critical to acquire an in-depth understanding of the interrelationships between structural vibrations and sound radiation. All these methods are discussed in this paper, their advantages and limitations are compared, and the need for further development to analyze the root causes of noise and vibration problems is discussed.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977731]

PACS number(s): 43.60.Sx, 43.60.Pt, 43.60.Jn [ADP] Pages: 2680–2697

I. INTRODUCTION

The relationship between structural vibration and sound radiation has long been a subject of interest in the acoustics community. Here, vibration is regarded as the cause and sound as the effect. When the cause is specified and effect is sought, it is known as a prediction (forward) problem. When the effect is measured and cause is desired, it is called a reconstruction (inverse) problem. In general, the latter is more difficult to deal with than the former because input data in an inverse problem are often incomplete and inaccurate; therefore it is mathematically ill posed. All methods that deal with inverse problems must address this ill-posedness difficulty. The focus of the present paper is to give an overview of the methods developed over the past three decades for visualizing the acoustic quantities in three-dimensional (3D) space and on source surfaces, thus the name of acoustic imaging methods, based on the acoustic pressure signals measured at very close distances around a target source.

Note that the field of acoustic imaging is very broad and its applications may be found in medical ultrasonic diagnosis, nondestructive evaluations and testing, sound navigation imaging or sonar, and underwater acoustic imaging, just to name a few. The present paper only focuses on a very specific topic of reconstruction of the acoustic quantities on a source surface, given the acoustic pressures measured in a near field of the source surface. Hence, many methods such as inverse frequency response function methods, inverse finite element methods, beamforming, and time reversal are all excluded, even though they fall in the category of inverse acoustic problems.

Direct applications of the acoustic imaging methods that the present paper is focused on may be found in noise diagnosis, where the root causes of a vibrating structure such as an engine or automobile are sought. Traditionally, noise diagnosis is carried out by using a microphone to measure the sound pressure level (SPL) and spectrum of a target structure to determine the noise levels and frequency contents. Another common approach is to scan a sound intensity probe over a vibrating structure to identify the "hot spot" that may indicate a sound source and the direction in which the acoustic energy is flowing out of this sound source.

The main advantage of these traditional methods is that they provide direct measurements of specific acoustic quantities such as SPL and sound intensity. Their limitations are that: (1) they provide specific acoustic quantities at the measurement location only; (2) the measured values are discrete and uncorrelated; therefore, it is not possible to get a global view of a sound field; (3) it is difficult to pinpoint the location of a noise source, which is especially true when there are other sources or reflecting surfaces nearby; and (4) it is not possible to identify the structural waves that are traveling along the surface or to visualize the out-of-plane vibration pattern, which may have a direct impact on the resultant structure-borne sound.

These limitations can be circumvented by using the methods to be surveyed in this paper. The major advantage of these methods is that it enables one to reconstruct all acoustic quantities such as the acoustic pressure, particle velocity, and acoustic intensity not only at a measurement location but in 3D space and on a source surface by measuring the acoustic pressure in the near field of the target source surface. Moreover, they allow for visualization of the structural waves traveling along the surface of a structure, yielding an invaluable insight into the interrelationship between sound and vibrations.

 © 2008 Acoustical Society of America

By reconstructing a 3D acoustic pressure field, we can acquire the insight of the radiation pattern, sound pressure distribution, and locations of airborne sources. Reconstructing the normal surface velocity distribution allows for visualization of the operational deflection shape of a complex structure, and reconstructing the acoustic intensity vector field allows for visualization of not only sound sources but also the flow of acoustic energy from sources to surrounding fluid medium. The insight provided by the acoustic imaging methods cannot be matched by traditional measurement methodologies.

In this paper, we review the development of these methods over the past three decades. In particular, we discuss some practical methods that are suitable for engineering applications. Note that the articles cited here are from refereed journals only. The conference papers and other nonrefereed articles are excluded. This is the present author's preference since most of the important conference papers are subsequently revised and published in refereed journals. Finally, we outline the prospects of the acoustic imaging methods for future engineering applications.

II. NEAR-FIELD ACOUSTICAL HOLOGRAPHY (NAH)

It is fair to say that NAH technology has fundamentally changed noise diagnostics in that it enables one to get all the acoustic quantities, including the acoustic pressure, particle velocity, acoustic intensity vector, sound power, and an out-of-plane structural vibration pattern, simply by taking the acoustic pressure measurements at very close distances to a target source. The insights into the acoustic characteristics of a sound source that one can get from NAH cannot be matched by any traditional measurement methods.

Note that NAH is evolved from acoustical holography[1] that utilizes the same principles as those of laser holography and visualizes the acoustic sources based on far-field acoustic pressure measurements.[2] Since the evanescent waves that decay exponentially with distance are lost in the far-field measurement, spatial resolution of acoustical holography is limited to the wavelength of an acoustic wave of interest.[3] In other words, one will not be able to discern acoustic sources that are separated by a distance less than a wavelength. This is not a problem for laser holography, as the wavelength of a leaser beam is in the nanometer range, so its spatial resolution is very high.

To enhance the spatial resolution to well within an acoustic wavelength, it is necessary to take measurements in the near field so evanescent waves can be included in the back propagation process. Theoretically, NAH can yield an unlimited spatial resolution in reconstruction, provided that measurement is taken continuously in the near field and all evanescent waves are captured.

The original NAH relies on Fourier transform technique that enables one to transform the acoustic pressure from temporal frequency domain to spatial frequency domain, project it from a hologram surface to any surface in a source-free region, and take an inverse Fourier transform to the temporal frequency domain. This approach is suitable for a structure that can be described by separable coordinate systems. For detailed discussions of using Fourier transform to reconstruct the desired acoustic quantities in 3D space and on the surface with separable coordinates, readers are referred to the research papers.[4–8]

In practice, this Fourier transform based NAH can be conducted in steps as follows. First, one selects a coordinate system for a given source structure. If this structure is a plate or consists of planar surfaces, the rectangular coordinate system is preferable. If a structure has a cylindrical shape, the cylindrical coordinate system is ideal. If it is spherical, the spherical coordinate system should be used. In any event, the source surface should be describable by separable coordinates.

Once a coordinate system is selected, one can set up a measurement grid conformal to the source surface at very close distances. For planar NAH, the measurement aperture should be four times as large as the source surface to ensure that sufficient information is captured. For example, for a plate of width W and length L, the measurement aperture should be at least $4\times(W\times L)$.[9,10] Measurement distance should be kept as close to the source surface as possible. Note that spatial resolution of the reconstructed acoustic quantity is critically dependent on measurement distance and signal to noise (S/N) ratio. For example, if the desired spatial resolution in the x-axis direction is $R_x=\lambda_x/2$, where λ_x is the wavelength of the shortest structural wave in the x-axis direction, $S/N=D$, then the measurement distance should be less than $d<(R_xD)/(20\pi\log e)$.[10] The higher the spatial resolution requirement, namely, the smaller the value of R_x value is, and the lower the S/N ratio D becomes, the closer the measurement aperture to the source surface d should be.

Note that the microphone spacing should be equal to or less than the measurement distance. This is because large microphone spacing will result in an insufficient spatial sampling and cause aliasing in reconstruction. Once the microphone spacing and measurement size are specified, the total measurement points are determined. For example, consider reconstructing the acoustic field in 3D space and on the source surface that is produced by a vibrating plate of dimension $(W\times L)$. Suppose that the S/N ratio is D and the desired spatial resolution is R. Then measurement distance is $d<(R\times D)/(20\pi\log e)$. If we set δ as microphone spacing, then the measurement points needed are $N=4\times(W\times L)/\delta^2$. To put this in numbers, let us consider a rectangular plate of dimensions $W\times L=0.4\times0.6$ m^2. Assume that S/N ratio $D=40$ dB, the required spatial resolution $R=0.05$ m, and $\lambda_{\min}=2R$, then the measurement distance must be $d<(0.05\times40)/(20\pi\log e)=0.073$ m. Let us take $d=0.07$ m and set microphone spacing at $\delta=\lambda_{\min}/3=0.1/3=0.033$ m. The total number of measurement points required becomes $N=4\times0.24/0.033^2=882$. To achieve a higher spatial resolution in reconstruction, the acoustic pressure must be measured at distances closer than 0.07 m and the microphone spacing must be smaller than 0.033 m. This example shows how to set up the measurement array and specify the measurement parameters, given the S/N ratio, required spatial resolution, and microphone spacing.

TABLE I. Fourier transform based NAH for planar, cylindrical, and spherical geometries.

Planar NAH	$p(x,y,z_S;\omega)=\mathcal{F}_x^{-1}\mathcal{F}_y^{-1}\{\mathcal{F}_x\mathcal{F}_y[p(x,y,z_h;\omega)]G_p(k_x,k_y,z_S-z_h)\}$, $v_n(x,y,z_S;\omega)=\mathcal{F}_x^{-1}\mathcal{F}_y^{-1}\{\mathcal{F}_x\mathcal{F}_y[p(x,y,z_h;\omega)]G_v(k_x,k_y,z_S-z_h)\}$, where $G_p(k_x,k_y,z_S-z_h)=e^{ik_z(z_h-z_S)}$, $G_v(k_x,k_y,z_S-z_h)=\frac{k_z}{\rho_0 ck}e^{ik_z(z_h-z_S)}$, $\mathcal{F}_x\mathcal{F}_y[\varphi]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi e^{-ik_x x}e^{-ik_y y}dxdy$, $\mathcal{F}_x^{-1}\mathcal{F}_y^{-1}[\Phi]=\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi e^{ik_x x}e^{ik_y y}dk_x dk_y$, and $k_z=\sqrt{k^2-k_x^2-k_y^2}$.
Cylindrical NAH	$p(r_S,\phi,z;\omega)=\mathcal{F}_\phi^{-1}\mathcal{F}_z^{-1}\{\mathcal{F}_\phi\mathcal{F}_z[p(r_h,\phi,z;\omega)]G_p(k_\phi,k_z,r_S-r_h)\}$, $v_n(r_S,\phi,z;\omega)=\mathcal{F}_\phi^{-1}\mathcal{F}_z^{-1}\{\mathcal{F}_\phi\mathcal{F}_z[p(r_h,\phi,z;\omega)]G_v(k_\phi,k_z,r_S-r_h)\}$, where $G_p(k_\phi,k_z,r_S-r_h)=\frac{H_n^{(1)}(k_r r_S)}{H^{(1)}(k_r r_h)}$, $G_v(k_\phi,k_z,r_S-r_h)=\frac{k_r}{i\rho_0 ck}\frac{H_n'^{(1)}(k_r r_S)}{H_n^{(1)}(k_r r_h)}$, $\mathcal{F}_\phi\mathcal{F}_z[\varphi]=\frac{1}{2\pi}\int_0^{2\pi}\int_{-\infty}^{\infty}\varphi e^{-in\phi}e^{-ik_z z}d\phi dz$, $\mathcal{F}_\phi^{-1}\mathcal{F}_z^{-1}[\Phi]=\frac{1}{2\pi}\Sigma_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi e^{in\phi}e^{ik_z z}dk_z$, and $k_r=\sqrt{k^2-k_z^2}$.
Spherical NAH	$p(r_S,\theta,\phi;\omega)=\mathcal{F}_\theta^{-1}\mathcal{F}_\phi^{-1}\{\mathcal{F}_\theta\mathcal{F}_\phi[p(r_h,\theta,\phi;\omega)]G_p(k_\theta,k_\phi,r_S-r_h)\}$, $v_n(r_S,\theta,\phi;\omega)=\mathcal{F}_\theta^{-1}\mathcal{F}_\phi^{-1}\{\mathcal{F}_\theta\mathcal{F}_\phi[p(r_h,\theta,\phi;\omega)]G_v(k_\theta,k_\phi,r_S-r_h)\}$, where $G_p(k_\theta,k_\phi,r_S-r_h)=\frac{h_n^{(1)}(kr_S)}{h_n^{(1)}(kr_h)}$, $G_v(k_\theta,k_\phi,r_S-r_h)=\frac{h_n'^{(1)}(kr_S)}{i\rho_0 ch_n^{(1)}(kr_h)}$, $\mathcal{F}_\theta\mathcal{F}_\phi[\varphi]=\int_0^{2\pi}\int_0^{\pi}\varphi Y_n^m(\theta,\phi)^*\sin\theta d\theta d\phi$, and $\mathcal{F}_\theta^{-1}\mathcal{F}_\phi^{-1}[\Phi]$ $=\Sigma_{n=0}^{\infty}\Sigma_{m=-n}^{n}\Phi Y_n^m(\theta,\phi)$.

Similar formulations for determining the measurement setup under specific measurement conditions using the cylindrical and spherical coordinates can be found in Ref. 10.

If the source surface is not exactly planar, cylindrical, or spherical, this Fourier transform approach is still applicable, except that reconstruction is valid in a source-free region bounded by two parallel or concentric surfaces.

For an exterior problem using the rectangular coordinate system, one parallel surface can be tangential to the target source surface and another is at infinity. Note that this approach is not applicable to an interior problem because the source-free requirement may not be met. Similarly, for the cylindrical and spherical coordinate systems in an exterior region, one concentric surface can be the minimum surface that circumscribes the target source and another is at infinity. Once again, this approach is invalid for the cylindrical coordinate system inside an enclosed cylindrical surface because the source-free requirement is violated. However, this approach will be valid for the spherical coordinate system inside a cavity, provided that sources are outside the maximum sphere of radius $r=r_S$ that is tangential to the interior surface of the cavity. The region of validity in this case will be inside the maximum sphere, so there is no way of reconstructing the acoustic quantities on an arbitrary interior surface.

Once the measurement setup is determined, we can collect the acoustic pressures and take a fast Fourier transform with respect to time to obtain spectral information. Next we take another Fourier transform with respect to spatial variable and convert the acoustic pressure into the wave number domain. The key in NAH, however, is to project acoustic quantities from a measurement surface to a parallel or concentric surface in a source-free region. This is something that none of the traditional methods can do. While Fourier transform is performed in the traditional methods, the locations are always fixed at the measurement points. In contrast, NAH enables one to obtain all acoustic quantities in the entire source-free region. This special capability is epitomized in the following formulation for, say, planar NAH:

$$P(k_x,k_y,z) = P(k_x,k_y,z_h)e^{ik_z(z-z_h)}, \tag{1}$$

where $P(k_x,k_y,z_h)$ and $P(k_x,k_y,z)$ stand for the angular spectra of the acoustic pressures evaluated at the measurement and a parallel surface in a source-free region, respectively, k_x and k_y represent wave numbers of the structural waves in the x- and y-axis directions, respectively, and $e^{ik_z(z-z_h)}$ is known as the propagator for reconstructing the acoustic pressure, which is a phase shift from the measurement plane to any parallel plane.

Equation (1) shows that the angular spectrum at any parallel plane, regardless whether it is toward the source plane or away from it, can be obtained by multiplying the angular spectrum at a hologram plane by a phase shift between two planes. Once this is done, the acoustic pressure in the temporal frequency domain can be obtained by taking an inverse spatial Fourier transform,

$$\hat{p}(x,y,z) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}P(k_x,k_y,z)e^{i(k_x x+k_y y)}dk_x dk_y. \tag{2}$$

Substituting Eq. (1) into Eq. (2) yields (Table I gives solutions in other coordinate systems)

$$\hat{p}(x,y,z) = \mathcal{F}_x^{-1}\mathcal{F}_y^{-1}\{P(k_x,k_y,z_h)G_p(k_x,k_y,z-z_h)\}, \quad (3)$$

where $G_p(k_x,k_y,z-z_h)=e^{ik_z(z-z_h)}$ and $k_z=\sqrt{k^2-(k_x^2+k_y^2)}$ is a wave number in the direction of wave propagation. Note that when the structural wave number k_x or k_y is larger than the acoustic wave number k, or equivalently, the wavelength of a structural wave λ_x or λ_y is shorter than that of the acoustic wave λ, k_z is purely imaginary and the amplitude of the acoustic pressure decays exponentially with distance z, known as an evanescent wave. When k_x and k_y are smaller than k, k_z is real and a sound wave is generated. When $k=\sqrt{k_x^2+k_y^2}$, $k_z=0$, known as the coincidence frequency, at which the structural wave moves in sync with an acoustic wave and the radiation efficiency is maximum.[11]

When a structure vibrates at a constant frequency, the resultant acoustic wave will have a constant frequency. Meanwhile, the structural vibration can be decomposed into structural waves with different wave numbers or wavelengths. Some structural wave numbers are smaller than the acoustic wave number; or equivalently, some structural wavelengths are longer than the acoustic wavelength. These long wavelength structural waves will support sound radiation. The majority of the structural wave numbers, however, are larger than the acoustic wave number; or equivalently, the majority of the structural wavelengths are shorter than the acoustic wavelength, which are called the evanescent waves and cannot support sound radiation. This explains why sound can be generated by vibrations but not all vibrations can produce sound.

Note that one can replace z in Eq. (1) by, say, z_S to determine the acoustic pressure on the source surface or anywhere in the source-free region. Equation (1) is the essence of NAH, which enables one to project the acoustic quantities from one surface to another in a source-free region.

Using Euler's equation, we can reconstruct the particle velocity and acoustic intensity on any parallel surface,

$$\hat{v}(x,y,z) = \mathcal{F}_x^{-1}\mathcal{F}_y^{-1}\{P(k_x,k_y,z_h)G_v(k_x,k_y,z-z_h)\}, \quad (4)$$

$$I_{\mathrm{av}}(x,y,z) = \tfrac{1}{2}\,\mathrm{Re}[\hat{p}(x,y,z)\hat{v}^*(x,y,z)], \quad (5)$$

where $G_v(k_x,k_y,z-z_h)=(k_z/\rho_0 ck)e^{ik_z(z-z_h)}$ is the propagator for the particle velocity.

Equations (1)–(5) illustrate that once the acoustic pressures are measured in the near field of a target source surface, all the acoustic quantities in a source-free region can be reconstructed and structural waves traveling on a source surface be visualized. This is the power of NAH.

Note that a finite measurement surface, also known as an aperture, can cause wraparound errors in reconstruction. In other words, it can introduce artificial wave numbers that are actually nonexistent. The amount of wraparound errors cannot be determined exactly but the effect of the aperture can be greatly reduced by increasing the aperture size to four times as large as a source size.[9,10] Hayek and Luce[12] showed that it is possible to reduce the measurement aperture to the size of a source surface, yet still allowing reconstruction of the source field with sufficient accuracy. Recently, patch NAH is developed[13–16] that uses analytic continuation of the patch pressure and singular value decomposition (SVD) to eliminate the need to scan a large measurement aperture. This patch NAH enables one to significantly reduce the number of measurement points normally required in NAH in tackling a large-scale structure.[17]

Another aperture effect is induced by a sudden drop in the amplitude of acoustic pressure at the edge of a measurement aperture. This discontinuity exists no matter how large an aperture is. As a result, there is a high concentration of very high wave numbers in the spectrum that may pose a serious problem in reconstruction, because any small errors in the measured data that are associated with high wave numbers will be significantly amplified, thus distorting reconstructed images. This difficulty is alleviated by applying a spatial window function that brings down the amplitudes of acoustic pressures at edge to zero gradually,[18] known as a Tukey window,[19] which combines cosine and rectangle windows to reduce the amplitude of the acoustic pressure to zero at boundaries of a measurement aperture without reducing the gain of the windowed transform.

Planar NAH has been employed to reconstruct the structure-borne intensity as well as the normal acoustic intensity, a technique known as structural intensity from measurements of the acoustic pressure (SIMAP).[20] Williams *et al.*[20] showed that an acoustic intensity map can fail to find the root cause of structural vibration. To identify the real driver of structural vibration, one needs to rely on a structural intensity map. Further, one needs a supersonic intensity map to identify the locations and strengths of the actual acoustic sources of a structure.[21] This supersonic intensity is responsible for sound radiation to the far field, which can be obtained by integrating the acoustic intensity over wave numbers bounded by the radiation circle, where the acoustic wave number is exactly the same as the structural wave number.

Loyau and Pascal[22] presented a variation of NAH and showed that the acoustic field from a planar source can be reconstructed by measuring the acoustic intensity using an intensity probe. Williams *et al.*[23] showed that planar NAH was extendable to broadband excitations by defining a two-dimensional spatial Fourier transform of the measured acoustic pressure signals divided by a driving force. In particular, they demonstrated an example of broadband NAH via reconstruction of the normal velocities on a cylindrical surface.[23]

III. INVERSE BOUNDARY ELEMENT METHOD (IBEM)

The concept of NAH can be generalized by using Green's function that satisfies either the homogeneous Dirichlet or Neumann condition on a measurement plane.[9,24–27] Theoretically, an acoustic field can be reconstructed if Green's function is specified. Such Green's function is available for simple geometry, for example, an infinite plane, but extremely difficult to derive for arbitrarily shaped geometry. One alternative is to solve the Fredholm integral equation of the first kind by approximating an acoustic quantity as a linear combination of a set of functions with the associated coefficients determined by an implicit least-squares method (LSM).[28] Another approach is to employ the Helmholtz integral theory[29]

$$\hat{p}(\vec{x}) = \frac{1}{4\pi}\int\int_S \left[\hat{p}(\vec{x}_S)\left(ik - \frac{1}{R}\right)\frac{e^{ikR}}{R}\frac{\partial R}{\partial n} - i\omega\rho_0\hat{v}_n(\vec{x}_S)\frac{e^{ikR}}{R}\right]dS, \tag{6}$$

$$\hat{p}(\vec{x}_{S'}) = \frac{1}{2\pi}\int\int_S \left[\hat{p}(\vec{x}_S)\left(ik - \frac{1}{R_S}\right)\frac{e^{ikR_S}}{R_S}\frac{\partial R_S}{\partial n_{S'}} - i\omega\rho_0\hat{v}_n(\vec{x}_S)\frac{e^{ikR_S}}{R_S}\right]dS, \tag{7}$$

where $\vec{x}$ and $\vec{x}_S$ represent points in the field and on the source surface, respectively, e^{ikR}/R is the free-space Green's function, $R=|\vec{x}-\vec{x}_S|$ implies the distance between a field and surface points, $R_S=|\vec{x}_{S'}-\vec{x}_S|$ indicates the distance between two surface points, and ρ_0 is the ambient density of the fluid medium.

Equation (6) shows that the complex amplitude of the acoustic pressure at any field point $\hat{p}(\vec{x})$ can be specified by a surface integral of the surface acoustic pressure $\hat{p}(\vec{x}_S)$ and the normal surface velocity $\hat{v}_n(\vec{x}_S)$ through the free-space Green's function. Meanwhile, the surface acoustic pressure $\hat{p}(\vec{x}_S)$ is related to the normal surface velocity $\hat{v}_n(\vec{x}_S)$ via Eq. (7). In a forward problem, either $\hat{p}(\vec{x}_S)$ or $\hat{v}_n(\vec{x}_S)$ is given, and Eq. (7) is used to solve for the other. Once $\hat{p}(\vec{x}_S)$ and $\hat{v}_n(\vec{x}_S)$ are specified, we can predict the acoustic pressure anywhere in free space by using Eq. (6).

Here neither $\hat{p}(\vec{x}_S)$ nor $\hat{v}_n(\vec{x}_S)$ is known a priori. Instead, the field acoustic pressure $\hat{p}(\vec{x})$ is given through measurements. Our goal is to reconstruct $\hat{p}(\vec{x}_S)$ and $\hat{v}_n(\vec{x}_S)$ based on measured values of $\hat{p}(\vec{x})$. To this end we solve Eqs. (6) and (7) simultaneously. Since there are no analytic solutions to these integral formulations for an arbitrary surface, numerical solutions are sought.

The most popular numerical method for solving an integral equation is boundary element method (BEM).[30–33] In BEM, the surface of an arbitrary object is discretized into small elements and an interpolation scheme such as the second-order isoparameteric interpolation function, also known as the shape function, is utilized to approximate the surface acoustic pressure $\hat{p}(\vec{x}_S)$ and normal surface velocity $\hat{v}_n(\vec{x}_S)$ at all discrete nodes.

Discretizing Eqs. (6) and (7), and replacing $\hat{p}(\vec{x}_S)$ with $\hat{v}_n(\vec{x}_S)$ or $\hat{v}_n(\vec{x}_S)$ with $\hat{p}(\vec{x}_S)$ in Eq. (6) by using Eq. (7), we can correlate the measured pressure to the surface acoustic pressure $\hat{p}(\vec{x}_S)$ and normal surface velocity $\hat{v}_n(\vec{x}_S)$, respectively, through the following equations:

$$\mathbf{p}(\mathbf{x}_m) = \mathbf{T}_p(\mathbf{x}_m|\mathbf{x}_S)\mathbf{p}(\mathbf{x}_S) \quad \text{and} \qquad \mathbf{p}(\mathbf{x}_m) = \mathbf{T}_v(\mathbf{x}_m|\mathbf{x}_S)\mathbf{v}_n(\mathbf{x}_S), \tag{8}$$

where $\mathbf{p}(\mathbf{x}_m)$ and $\mathbf{p}(\mathbf{x}_S)$ represent two column vectors containing the acoustic pressures measured on a hologram surface at $\mathbf{x}_m$ and on a source surface at $\mathbf{x}_S$, respectively, $\mathbf{v}_n(\mathbf{x}_S)$ is a column vector containing the normal surface velocity at $\mathbf{x}_S$, $\mathbf{T}_p(\mathbf{x}_m|\mathbf{x}_S)$ and $\mathbf{T}_v(\mathbf{x}_m|\mathbf{x}_S)$ are the transfer matrices that correlate $\mathbf{p}(\mathbf{x}_m)$ to $\mathbf{p}(\mathbf{x}_S)$ and $\mathbf{v}_n(\mathbf{x}_S)$, respectively,

$$\mathbf{T}_p(\mathbf{x}_m|\mathbf{x}_S) = (4\pi)^{-1}\{\mathbf{D}(\mathbf{x}) + \mathbf{M}(\mathbf{x})\mathbf{M}_S^{-1}(\mathbf{x}_S)[2\pi\mathbf{I} - \mathbf{D}_S(\mathbf{x}_S)]\}, \tag{9a}$$

$$\mathbf{T}_v(\mathbf{x}_m|\mathbf{x}_S) = (4\pi)^{-1}\{\mathbf{D}(\mathbf{x})[2\pi\mathbf{I} - \mathbf{D}_S(\mathbf{x}_S)]^{-1}\mathbf{M}_S(\mathbf{x}_S) + \mathbf{M}(\mathbf{x})\}, \tag{9b}$$

where $\mathbf{M}_S(\mathbf{x}_S)$ and $\mathbf{D}_S(\mathbf{x}_S)$ represent the effects of the monopoles and dipoles on a surface point, respectively, while $\mathbf{M}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$ represent those of the monopoles and dipoles on a field point, respectively [see Appendix for the formulations of matrices $\mathbf{M}_S(\mathbf{x}_S)$, $\mathbf{D}_S(\mathbf{x}_S)$, $\mathbf{M}(\mathbf{x})$, and $\mathbf{D}(\mathbf{x})$].

The first attempt to reconstruct the acoustic quantities in an interior region bounded by an arbitrarily shaped body was shown by Gardner and Bernhard,[34] who utilized IBEM to display the interaction of the acoustic sources on the surface and field points in an interior region.

The formal derivations of the IBEM formulation were given by Veronesi and Maynard.[35] In particular, the acoustic pressure $\mathbf{p}(\mathbf{x}_S)$ and normal velocity $\mathbf{v}_n(\mathbf{x}_S)$ in Eq. (8) are solved using SVD in an inverse acoustic problem.[35]

An example of using SVD for an axisymmetric object was illustrated by Borgiotti *et al.*,[36] in which the measurements were taken on a surface conformal to the source geometry as close as feasible. Since Eq. (8) is an ill-posed problem, it is regularized by truncating matrices to include only the singular values that are larger than a certain tolerance.[37]

Over the next ten years many research papers emerged that used SVD and the Helmholtz integral formulations to reconstruct acoustic radiation from an arbitrary structure.[38–48] SVD is a powerful technique for solving a general matrix equation.[49] It enables one to express any $(M \times N)$ complex matrix as a diagonal matrix in proper bases together with the domain and range spaces. SVD yields the least-squares solution for an overdetermined system of equations $(M > N)$.[50]

The most valuable information that SVD offers is the singular values of ill-conditioned or singular matrices. When an ill-conditioned matrix is inverted, small singular values become huge numbers. Consequently, any error embedded in the small singular values can be greatly amplified and reconstructed images are distorted. By eliminating small singular values of an ill-conditioned matrix, the matrix equation may be stabilized to yield a well-conditioned solution. This truncated SVD (TSVD) will work for an ill-posed problem when other elementary methods fail. However, numerical computations involving SVD can be time consuming, especially at high frequencies.

To circumvent intensive numerical calculation in SVD, Sarkissian[51] developed a different algorithm for axisymmetric geometry, where the surface normal velocity is expanded in terms of a set of orthonormal functions that are independent of frequency. Such an algorithm enables one to separate the part of the field that is very sensitive and susceptible to noise contamination from that of the field that can be reconstructed. Sarkissian[52,53] also expanded the acoustic pressure and normal component of the particle velocity outside a radiating sphere using the outgoing spherical waves. Since ex-

pansion functions are orthogonal in the far field, this approach is suitable for far-field acoustic holography.

Norris[54] derived an asymptotic approximation of the Helmholtz integral formulation for a high-frequency approximation of far-field radiation pattern from a cylinder with a nonseparable cross section. Villot *et al.*[55] extended NAH to a semi-infinite rectangular duct, where the radiator was placed at one end, and labeled it phonoscopy. The acoustic field inside this duct is expanded in terms of the normal modes over the duct cross section and one-dimensional plane waves in the longitudinal direction. Measurements are taken on two closely spaced parallel surfaces.

The advantages of IBEM are as follows: (1) it allows for reconstruction of the acoustic quantities on an arbitrarily shaped structure; (2) it has no restrictions on the locations of measurement points on a hologram surface, so long as they are in a near field; (3) it has no restrictions on the locations of reconstruction points, whether they are on an arbitrary source surface or in a free field; and (4) it is suitable for exterior and interior regions. However, the acoustic quantities are obtained through a spatial discretization, so one must require a minimum number of nodes per wavelength, typically 2π or 6, to ensure a satisfactory spatial resolution in reconstruction. Each node will contain two unknown variables: the surface acoustic pressure and the normal surface velocity. For a complex structure such as an automobile, the number of discrete nodes required to accurately reconstruct the surface acoustic quantities can be very large. Accordingly, the total number of measurement points may be excessive, making the reconstruction process extremely time consuming, which is especially true in the midfrequency range and above.

As an example, let us consider the same vibrating plate as that cited in Sec. II. For consistence, let us set the shortest wavelength at twice the desired spatial resolution, $\lambda_{\min}=2R=0.1$ m. The microphone spacing is set at $\delta=\lambda_{\min}/6=0.017$ m, as it is usually done in generating a BEM mesh, and measurement distance is equal to microphone spacing, $d=0.017$ m. Suppose that the measurement aperture is the same size as that of a source $S=0.24$ m^2 on one side of the plate only. The total measurement points needed are found to be $N=S/\delta^2=0.24/0.017^2=864$. Considering the size of the measurement aperture, this measurement number is very high. This is because one needs roughly six nodes per wavelength of interest in BEM computations.

This example indicates that although IBEM is based on rigorous integral formulations, its implementation is a challenge because it requires an excessive number of measurement points.

IV. LEAST-SQUARES METHOD (LSM)

An alternative to the Fourier transform and IBEM approaches is the LSM.[56,57] Unlike the first two approaches, LSM does not seek an exact solution to the acoustic field generated by an arbitrary surface, but rather an approximate one for an acoustic field using an expansion of certain admissible basis functions with its errors minimized by least squares. As such, LSM enables one to tackle complex situations with relatively few measurements.

Note that one can choose different coordinate systems for LSM expansions. The spherical coordinates offer approximate solutions in terms of the spherical wave functions. The cylindrical coordinates give approximate solutions in the cylindrical wave functions. The coefficients in the expansions are specified by matching the assumed-form solution to measured acoustic pressures, and the errors are minimized by least squares.

Mathematically, LSM formulations are expressible in matrix equations as[57]

$$\mathbf{p}(\mathbf{x}) = \mathbf{G}_p(\mathbf{x}|\mathbf{x}_m)\mathbf{p}(\mathbf{x}_m) \quad \text{and} \quad \mathbf{v}(\mathbf{x}) = \mathbf{G}_v(\mathbf{x}|\mathbf{x}_m)\mathbf{p}(\mathbf{x}_m), \tag{10}$$

where $\mathbf{p}(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$ represent the column vectors of the acoustic pressure and particle velocity at any desired location $\mathbf{x}$, respectively, which can be in the field or on a source surface, $\mathbf{p}(\mathbf{x}_m)$ is the column vector that contains the acoustic pressure measured on a hologram surface, $\mathbf{G}_p(\mathbf{x}|\mathbf{x}_m)$ and $\mathbf{G}_v(\mathbf{x}|\mathbf{x}_m)$ stand for the transfer matrices that correlate the measured pressure $\mathbf{p}(\mathbf{x}_m)$ to $\mathbf{p}(\mathbf{x})$ or $\mathbf{v}(\mathbf{x})$ at any field point including the source surface, respectively,

$$\mathbf{G}_p(\mathbf{x}|\mathbf{x}_m) = \boldsymbol{\Psi}(\mathbf{x})\boldsymbol{\Psi}(\mathbf{x}_m)^{\dagger} \quad \text{and}$$
$$\mathbf{G}_v(\mathbf{x}|\mathbf{x}_m) = \frac{1}{i\omega\rho_0}\nabla\,\boldsymbol{\Psi}(\mathbf{x})\boldsymbol{\Psi}(\mathbf{x}_m)^{\dagger}, \tag{11}$$

where $\boldsymbol{\Psi}(\mathbf{x}_m)^{\dagger}$ is known as a pseudoinversion,

$$\boldsymbol{\Psi}(\mathbf{x}_m)^{\dagger} = [\boldsymbol{\Psi}(\mathbf{x}_m)^H\boldsymbol{\Psi}(\mathbf{x}_m)]^{-1}\boldsymbol{\Psi}(\mathbf{x}_m)^H, \tag{12}$$

where the superscript H indicates a conjugate transposition of $\boldsymbol{\Psi}$, whose elements consist of the particular solution to the Helmholtz equation, which are expressible in spherical coordinates as

$$\Psi_j(r,\theta,\phi;\omega) \equiv \Psi_{nl}(r,\theta,\phi;\omega) = h_n^{(1)}(kr)Y_n^l(\theta,\phi), \tag{13}$$

where $h_n^{(1)}(kr)$ and $Y_n^l(\theta,\phi)$ are the spherical Hankel functions of the first kind and the spherical harmonics, respectively, and the indices j, n, and l in Eq. (13) are related via $j=n^2+n+l+1$ with n starting from 0 to N and l from $-n$ to $+n$.

Equation (10) gives an approximate solution to an acoustic field based on the information collected on a hologram surface at $\mathbf{x}_m$. LSM always attempts to find the best result for any given set of input data using optimization and least-squares processes. The more accurate the input data are, the better the approximation becomes. To this end, a conformal array of microphones should be used and placed at very close distances to a target source surface.

The major advantages of LSM are its simplicity in formulation, efficiency in computation, and flexibility in application. Since LSM solves the Helmholtz equation directly, it is immune to the nonuniqueness difficulty inherent in BEM based Helmholtz integral formulations.[56]

The major disadvantage of LSM is due to the fact that there is no single set of coordinate system that can yield a

good approximation for all surface geometries. For example, the spherical coordinate system is good for a blunt and convex surface, but not for a highly elongated one; the cylindrical coordinate system is ideal for a slender body, but not for a flat surface. Moreover, for an arbitrary surface the accuracy in reconstruction by using LSM is lower than that of IBEM, because of the approximate nature of LSM formulations.

It is worth pointing out that since Ψ_{nl} represents the outgoing spherical wave, the solutions provided by Eq. (10) converge very quickly when measurement and reconstruction points are outside a minimum sphere that circumscribes an arbitrary surface under consideration. In fact, when the measured acoustic pressures are exact, the reconstructed acoustic pressure converges to the true value as the number of expansion terms $J \to \infty$.[58] Inside a minimum sphere, however, situation was not clear in the beginning. The questions one might ask include the following: How can the sound field on the surface of an arbitrary structure be described by the spherical waves? Is this a Rayleigh hypothesis in NAH that pushes a formulation beyond its region of validity?

At the turn of the last century, Rayleigh[59,60] used a series expansion of plane waves to depict an acoustic field scattered from one-dimensional gratings and periodic corrugated structures. This series converged in the region above the highest point of the surface corrugations. However, Rayleigh extended his series to the entire field, including surface corrugations without giving any justification, thus known as a Rayleigh hypothesis. This series expansion was tested for acoustic scattering from arbitrary surfaces and aroused controversies over the next 60 years. Examples included acoustic scattering from cylinders with arbitrary cross sections, and the results given by an expansion of cylindrical wave functions could be completely wrong.[61]

These controversies were eventually settled by Millar[62–64] who proved that the Rayleigh hypothesis was neither completely right nor completely wrong. The validity of a Rayleigh series is governed by the locations of the singularities of analytic continuation of the exterior scattered field across a scattering surface. For example, in the case of scattering from gratings and periodic corrugated structures, the Rayleigh series solution is valid if the singularities lie below the lowest point of a corrugated surface. If the singularities lie above the lowest point of a corrugated surface, the Rayleigh series solution will be valid in the region above the highest singularities only. If this series solution is extended below these singularities, the results will be completely wrong.

Similarly, in a two-dimensional acoustic scattering problem, a Rayleigh series in terms of the outgoing cylindrical waves is valid outside a circle of radius defined by the singularities, but invalid inside this circle. The singularities in analytic continuation are obtained by Millar[65] as the algebraic branch points of the Schwarz function[66,67] of a boundary curve. The actual singularities depend on the Cauchy data on boundaries. van den Berg and Fokkema[61] provided the conditions, which are also obtained independently by Millar,[65] for which the Rayleigh hypothesis is valid in scattering from an infinite cylinder of an arbitrary cross section.

In 3D space, the Rayleigh series can be expanded in terms of the spherical Hankel functions and spherical harmonics with their coefficients determined by the orthogonality properties of the spherical harmonics. This spherical expansion falls naturally under the Rayleigh hypothesis. Thus, its solution is valid outside the minimum sphere that circumscribes an arbitrary source but invalid inside the minimum sphere depending on the locations of singularities.

On the surface, LSM seems similar to a Rayleigh series in three dimensions since it uses the spherical Hankel functions and spherical harmonics to approximate acoustic fields. Naturally, one would ask the following: Will LSM be subject to the same restrictions as a Rayleigh series does such that Eq. (10) is valid outside the minimum sphere that circumscribes an arbitrary object in an exterior region only? More fundamentally, one might ask the following: How is LSM different from a Rayleigh series in this case?

The interrelationships between LSM and Rayleigh series were revealed by Semenova and Wu[68] in reconstructing the acoustic fields generated by an arbitrary surface in the exterior region. In particular, they considered infinite cylinders with arbitrary cross sections and discovered that outside the minimum circle that circumscribed the singularities of a cylinder, the Rayleigh series yielded identical results as LSM did. This is because the high-order terms in the Rayleigh series are negligibly small so the differences between the Rayleigh series solutions and LSM solutions (a truncated expansion) are minuscule.

Of particular interest is what is inside a minimum circle. It is demonstrated that a Rayleigh series diverges inside a minimum circle bounded by singularities,[68] which confirms Millar's[62–64] theory on the validity of the Rayleigh hypothesis. On the other hand, LSM can yield reasonable results on an arbitrary surface without being subjected to the same restriction as a Rayleigh series. Of course, the further the reconstruction point goes inside a minimum circle, the larger the reconstruction errors become. It is interesting to point out, however, that even if the Rayleigh series is truncated at the same order as that of LSM, the errors in reconstruction inside the minimum circle are still much larger than those of LSM.[68]

These results suggest that LSM is different from a Rayleigh series as far as reconstruction of acoustic quantities on an arbitrary surface is concerned. A major difference lies in the fact that the former expresses the acoustic pressure in terms of an optimal number of expansion functions and uses least squares to determine expansion coefficients, while the latter describes the acoustic pressure as an infinite series and employs the orthogonality properties of the spherical harmonics to calculate the expansion coefficients. Nevertheless, this difference cannot be used to justify the validity of LSM inside a minimum sphere. Moreover, the previous results have depicted that the accuracy of reconstruction on a nonspherical surface using LSM decreases with the increase in the aspect ratio of a source surface and dimensionless frequency ka.[69]

A rigorous mathematical justification of LSM for reconstructing acoustic quantities on an arbitrary surface was provided by Isakov and Wu,[70] who proved that any radiating

solution to the Helmholtz equation outside a bounded Lipschitz domain with a connected complement could be approximated by a family of special solutions. Using this approximation and conditional stability estimates in the Cauchy problem for an elliptic equation, they demonstrated that the solution was bounded outside an arbitrarily shaped surface and converged to the exact solution, provided that it converged to the exact solution on the measurement surface. Further, they gave the estimate of convergence of Hölder and logarithmic type in different regions.[70] So they answered the question of the validity of LSM, namely, it is indeed possible to use the spherical wave functions to depict an acoustic field produced by a nonspherical surface.

Another impact of this paper is the suggestion of an effective regularization to overcome the ill-posedness difficulty inherent in an inverse acoustic problem.[70] Specifically, they proposed the regularization technique using a quasisolution,[71] i.e., an assumed-form solution, by imposing an a priori bound on the growth of the reconstructed acoustic pressure in an entire exterior region including a source surface S, which may be correlated to the acoustic power[72] and independent of the measurement locations. Using this constraint on a source surface, reconstruction errors were found to be finite everywhere including a source surface, whereas reconstruction errors using an iteration scheme alone grew to an unsatisfactory level as the acoustic quantities were propagated back toward a source surface.[68] This explains why the accuracy of reconstruction on the surface of an arbitrary structure using LSM with an iteration scheme may be unsatisfactory.[69]

The validity of using spherical wave functions and spherical harmonics to reconstruct the acoustic quantities on a nonspherical surface was investigated by Prager[73] as well. In particular, Prager proposed a method to approximate the sound field not fulfilling the Rayleigh hypothesis by transforming the nonconverging spherical wave function expansion to a converging one.

Like IBEM, LSM is executed through a conformal array of microphones placed in a near field of a vibrating structure. Unlike IBEM, LSM allows for patch reconstruction. For example, it enables one to reconstruct acoustic quantities on a portion of a source surface, which can be very convenient in practice. Specifically, one can take measurements over the area that is one row and one column of microphones larger than the surface area of interest. Since LSM can use the same criterion to set microphone spacing as that of Fourier transform based NAH, but covers a smaller area than NAH does, the measurement points for LSM are fewer than those for NAH.

As an example, let us consider the same plate as that in Sect. II with $S=0.24\ \mathrm{m}^2$. Assume that the shortest structural wavelength of interest is $\lambda_{\min}=0.1$ m, so we set $\delta=\lambda_{\min}/3=0.033$ m, which is the same as that in Fourier transform based NAH. The number of measurement points is $N=S/\delta^2=0.24/0.033^2=220$. Note that this is an estimate. The exact value of N is slightly higher because the measurement aperture is one row and one column of microphone grid larger than the surface area. In any event, this number is less than that of planar NAH, which is 882 because the measurement aperture is four times as large as that in LSM; and less than that of IBEM, which is 864 because the microphone spacing is smaller than that in LSM.

To ensure a satisfactory reconstruction by using LSM, it is a good idea to further decrease microphone spacing to, say, $\lambda_{\min}/4$ and measurement distance to, say, $d=\lambda_{\min}/8$. This will ensure that enough evanescent waves are captured so that spatial resolution in the reconstructed acoustic image for an arbitrary surface is enhanced.

It is important to point out that no matter how small the microphone spacing is and how close the measurement distance is to a source surface, the accuracy in reconstruction using LSM for an arbitrarily shaped surface will not be as high as that using IBEM when the measurement distance and microphone spacing are correctly selected. Likewise, for a source surface that is describable by separate coordinates so that Fourier transform is applicable, the reconstruction accuracy using LSM will not be as high as that using Fourier transform based NAH when the measurement area, distance, and microphone spacing are correctly specified. This is because LSM formulations are approximate, whereas IBEM and Fourier transform based formulations are exact.

On the other hand, IBEM requires an excessive number of measurement points, rendering this approach impractical for engineering applications; whereas Fourier transform based NAH is applicable for separable coordinates only, making it impossible to reconstruct acoustic quantities on an arbitrary surface in practice.

With LSM one can at least get an approximate solution to an acoustic field produced by a vibrating structure with relatively few measurement points. The accuracy in reconstruction using LSM depends on that in the input data. The more accurate the input data are, the more accurate the reconstruction becomes. LSM has been applied to reconstructing the acoustic fields produced by an arbitrary structure in both exterior[74,75] and interior regions.[76,77] Results demonstrate that LSM can provide satisfactory reconstruction with relatively fewer measurement points. In particular, it allows for a piecewise reconstruction, which can be very handy in engineering applications.

V. ADVANCEMENT OF TECHNOLOGY

In IBEM, it is important to ensure a linear independence of the transfer matrix by placing a measurement point in a proper position. The same problem has been studied extensively for an on-orbit modal identification of a large space structure. An effective independence (EfI) method has been developed to determine the optimum sensor locations from a large set of measurement points in order to maintain linear independence of target modal partitions.[78–80]

Kim and Ih[81] adopted EfI in NAH to identify the optimum measurement locations from a large number of measurement points in reconstructing a vibroacoustic field in an interior region. The contributions of various measurement sensors to a linear independence of the transfer matrix at any particular frequency were calculated, and linearly dependent points were discarded. This process was repeated until all sensor locations were least interdependent to each other for a

given frequency. In addition, SVD was used to obtain the least-squares solution and regularization was employed to stabilize the reconstructed acoustic quantities. In Ref. 82, EfI was used to determine the optimal measurement locations by eliminating all interdependent points through an iteration procedure. Moreover, an optimal wave-vector filter was used to suppress the high wave numbers that could distort the reconstructed acoustic image.

Errors in specifying sensor locations may induce high wave numbers that are susceptible to errors in reconstruction.[83,84] Adding low-pass filters may not be enough to beat these errors. A better approach is to apply an optimized Wiener filter[85] to enhance a S/N ratio, which is effective when background noise is independent of the spatial wave number. Fleischer and Axelrad[86] and Bai[87] demonstrated numerical examples of using this Wiener filter for planar structures.

One can supplement input data in reconstruction by measuring both the acoustic pressure and normal surface velocity. The concept of using both the acoustic pressure and normal surface velocity was examined numerically by Kang and Ih[88] on a rectangular plate. In this case the field acoustic pressures were measured at M points, while the normal surface velocities were specified on n $(n<M)$ surface nodes to form an overdetermined system of equations. The surface acoustic pressures and normal velocities were obtained by using the LSM and SVD.

Another attempt to improve the accuracy and efficiency in reconstruction was to combine LSM and IBEM algorithms, known as Combined HELS or CHELS for arbitrary surfaces.[89] The first step in CHELS is to measure the acoustic pressures at M points $(M=M_1+M_2)$ on a hypothetical spherical surface enclosing a structure of interest. A part of the input data M_1 are utilized to specify the expansion coefficients C_j, and the remaining data M_2 are used to calculate an optimal number of expansion term J_{op}. Once J_{op} is specified, additional field acoustic pressures are generated at as many points on a hologram surface as necessary to supplement the original data in reconstructing the acoustic quantities. The efficiency in reconstruction is enhanced because the acoustic pressure and normal surface velocity on N discrete nodes of an arbitrary surface are reconstructed by M measurement points. Since M is less than N, the measurement time and effort are reduced and the efficiency in reconstruction is enhanced.[89]

Note that in CHELS the additional field acoustic pressure will have the same accuracy as the measured data, since on the minimum sphere that circumscribes an arbitrarily shaped surface, the acoustic pressure given by Eq. (10) is uniformly convergent as $J\rightarrow\infty$. Therefore, there is no need to take more measurement than necessary.

Thus far, the reconstruction methods are developed for visualizing acoustic radiation in a free field. This free-field assumption may be true inside a fully anechoic chamber, but not true in a real environment because in reality space is often bounded by reflecting surfaces. Accordingly, the field acoustic pressure consists of direct and reflected acoustic waves, both of which must be accounted for in reconstruction.

To this end, a hybrid approach is developed that expresses the acoustic pressure in terms of outgoing and incoming spherical waves simultaneously.[90] These expansions are coupled with the BEM code to enhance the efficiency in reconstructing vibroacoustic responses on the surface of an arbitrary structure. Examples of this hybrid approach are shown in reconstruction of the vibroacoustic responses of a vibrating engine block[91] and that from an elongated circular cylinder with a diameter to length aspect ratio of 1:10.[92] The results are compared with those obtained by using IBEM alone, and good agreements are obtained in both cases.

It should be pointed out that since the acoustic pressures measured on a hologram surface contain the direct sound radiation from a source and sound reflections from nearby surfaces, and since the hologram surface encloses the source only but not reflecting surfaces, the reconstructed acoustic quantities may be satisfactory around the hologram and source surfaces only.

VI. TRANSIENT CASE

In practice, most vibrating structures are subject to transient excitations. Though there are numerous research papers on prediction of transient acoustic radiation from finite objects subject to transient excitations, papers on reconstruction of a transient acoustic field generated by a finite structure are very scarce. The closest work available is the study of an exchange of vibration and acoustic energies in a vibrating structure in the time domain through NAH measurement,[93] using near-field scanning over a spherical surface enclosing a sound source to predict transient acoustic radiation in the far field,[94] and a projection of transient acoustic pressures into a far field.[95] This scarceness is due to a lack of effective methodologies for reconstructing transient acoustic fields based on a finite number of the acoustic pressure measurements.

Wu *et al.*[96] are the first to use LSM to display visualization of transient acoustic radiation from an impulsively expanding sphere, an impulsively accelerated sphere, and a baffled sphere. The reason for selecting a spherical object is that analytic solutions are readily available so the results can be examined vigorously. Input data in these studies are produced by analytic formula,[97–100] which are evaluated on an arbitrarily selected line segment at arbitrary distances. To facilitate the derivations of a temporal solution, Laplace transform is utilized. The temporal kernel is obtained by replacing the spherical Hankel functions and their derivatives with polynomials, and the infinite integral in taking inverse Laplace transform is recast as a contour integral and evaluated by using the residue theorem.[10] The transient acoustic field is obtained by convoluting the temporal kernel with respect to the measured acoustic pressures.

VII. SOURCE IN MOTION

Another advancement is its extension to reconstruction of a coherent band limited sound field resulting from a moving source, known as moving frame acoustic holography (MFAH).[101] This MFAH was originally developed to increase an aperture size and reduce microphone spacing so as to obtain a better hologram. For stationary sources, this was

done by scanning the microphone array continuously over a measurement plane. MFAH was adopted in reconstructing the acoustic field produced by a moving source, in which a microphone array was fixed while the source was moving at a constant speed.

Since there is relative motion between a microphone array and source in MFAH, Doppler shift occurs in the measured acoustic pressure signals. As a result, the frequency band centered at any frequency of interest in reconstruction is broadened. This broadening effect increases with a relative speed between a microphone array and source in motion. So MFAH is inherently limited to reconstructing tonal sounds whose frequencies are so far apart that sideband overlapping does not occur.

Park and Kim[102] improved MFAH to make it suitable for a sound field with band-limited spectra. They assumed that the acoustic field generated by band-limited and coherent sources had the same spatial distribution for acoustic pressures at any frequency within the bandwidth. So the Doppler shifted spectrum may be regarded as convolution of two spectra, one being a normalized source spectrum and the other being the mean wave number spectrum in the direction of motion. Once the normalized source spectrum was measured, the mean wave number spectrum could be obtained by deconvolution in the time domain.

Another side effect of MFAH is that the source strength may be amplified in the front but suppressed in the back portion of the source surface.[103] This distortion is undesirable and must be corrected. Park and Kim[104] examined the case where the source moved rectilinearly at low speed (much less than the speed of sound in a fluid medium). They assumed that the acoustic pressures were measured in a far field and the acoustic fields were approximated by those from monopoles distributed on a surface in motion. These assumptions significantly simplify the problem because one can derive the amplification factor due to a moving monopole explicitly and correct the error due to relative motion between measurement microphones and the source. Also, Park and Kim[105] used MFAH to visualize vehicle pass-by noise, assuming that the acoustic field is semistationary. This assumption allows one to ignore the vehicle speed and frequency changes during the test.

Ruhala and Swanson[106] extended NAH to include the effect of a moving medium. In this study the microphone array was moving with a source at the same speed, so there was no relative motion between microphones and the source but there was mean flow between them. To account for the mean flow effect, they employed a special Green's function and wave number filters, and validated their results through a point source in a moving medium.

VIII. EQUIVALENT SOURCE METHOD

In 1995, Ochmann[107] proposed the idea of using a multipoint multipole method instead of a one-point multipole method to simulate acoustic radiation from arbitrary sources. Jeon and Ih[108] explored a similar idea in an effort to reduce the measurement points required in IBEM. In particular, they developed an equivalent source method where the field acoustic pressure was reproduced using point sources distributed inside the source surface. Also, they reformulated LSM formulations[57] by expanding the spherical Hankel functions and spherical harmonics with respect to points distributed in an interior region of a source surface. Contributions from equivalent sources were evaluated by matching the assumed-form solution to boundary conditions specified on the source surface[109–111] or to the acoustic pressures measured on a hologram surface.[57] The locations of the equivalent sources were optimized by a natural algorithm[112] or the EfI method.[74] The optimal number of expansion terms was determined by a spatial filter and a regularization scheme. Once the expansion coefficients were determined, the field acoustic pressure was recalculated to supplement the original input data to IBEM codes, just like CHELS algorithms did.[89]

Semenova and Wu[113] offered more general formulations by expanding the acoustic pressure with respect to multiple sources distributed inside the source surface. In particular, they examined the performances of three expansions: localized spherical waves (LSWs),[114] distributed spherical waves (DSWs),[115] and distributed point sources (DPSs),[116] for reconstruction of acoustic radiation from an arbitrary source by using the same input data. The reconstructed acoustic pressures were validated against benchmark values measured at the same locations as the reconstruction points for frequencies up to 3275 Hz. Reconstruction was obtained using Tikhonov regularization (TR) or its modification with regularization parameters selected by an error-free parameter choice method. These results revealed that DSWs provided a better-conditioned transfer matrix and more accurate reconstruction than LSWs and DPSs did, and were not affected as much by change in the number of measurement points.

IX. PARTIAL FIELD DECOMPOSITION

Recently, much effort has been made to visualize and analyze relative contributions from mutually incoherent sources.[117,118] In these studies, a number of reference microphones were placed next to the apparent sources, and individual contributions from these sources were calculated by taking cross correlations. Noise rejection method was used to discriminate relative contributions from individual sources.[119,120] Specifically, the effect of a target source on the overall signal was treated as a true signal and those from other sources as noise. By calculating coherence functions among all sensors and comparing them with that of the sensor close to a target source, one could recognize the contributions from this target source. The shortcoming of this approach was that it required a priori knowledge of the locations of individual sources.

To analyze contributions from individual sources without any prior information on source locations, Nam and Kim[121] proposed to utilize spatial information of a hologram to infer the source position. They introduced an optimal decomposition plane in order to reduce spatial overlap in reconstruction. This technique was based on a vector representation of an acoustic field together with the knowledge of the point-to-point transfer functions that were obtained by measurements and estimation through holographic projection.[122]

This method seemed to work well for a planar surface. For an arbitrary surface, however, a more effective method should be developed.

Kwon *et al.*[123] used multireference scan-based NAH to decompose the sound field into a number of mutually incoherent partial fields. These partial fields were then added quadratically to give the sound power and projected to any reconstruction surface. In carrying out partial field decomposition, the reference signals that were linearly proportional to a source signal were used instead of the source signal because in these cases source signals could not be measured directly. The number of reference signals that were fixed in space must be equal to or greater than that of incoherent sources so as to ensure successful partial field decomposition.

Lee *et al.*[124] made use of the partial field decomposition together with cylindrical NAH to study aeroacoustic sources. In particular, they developed procedures to determine the number of reference microphones needed to decompose measured acoustic pressure into partially correlated acoustic fields on a hologram surface. Moreover, they used source nonstationarity compensation and regularization to visualize partially correlated sound fields produced by ducted fans. Lee and Bolton[125] showed that noise in the reference signals and variations in a source level during scan-based measurements could be properly accounted for in the partial field decomposition.

Kim *et al.*[126] showed another study on using partial field decomposition for visualizing an acoustic field. The key in this partial field decomposition was to position reference microphones as close to incoherent sources as possible. However, the optimal reference microphone locations were unknown a priori. To overcome this difficulty, they developed postprocessing procedures that could identify the optimal reference microphone locations and then placed virtual references at these locations after acoustic pressures were measured on a hologram surface.

Nam and Kim[127] further extended the decomposition method they presented earlier[121] and offered partial field decomposition algorithms that did not require placing reference sensors close to individual sources. They used calculations rather than measurement to decompose a hologram image into images of incoherent sources. This approach simplified the holographic measurement processes. The steps involved in this approach included: (1) measuring the acoustic pressures on a hologram plane, (2) estimating the acoustic pressure on a source plane, (3) selecting the maximum acoustic pressure, (4) estimating the acoustic field due to one source, and (5) estimating the remaining acoustic field. These algorithms were validated in two experiments, and the results demonstrated that using this approach it was possible to identify hidden sources and estimate the acoustic power of each individual source.

X. RECENT APPLICATIONS AND DEVELOPMENTS

NAH has been utilized to analyze the real world noise and vibration problems. Examples of such were given by Williams *et al.*,[128] who used IBEM to analyze interior noise and fuselage vibrations of a propeller-driven aircraft Model Beech 1900 in-flight. The purpose of this project was to examine the feasibility of using NAH to analyze noise and vibration problems of a large structure under the actual operating conditions. This project shows that NAH can be extended to solving a wide range of industrial noise control problems.

Herdic *et al.*[129] illustrated vibroacoustic analyses of a full-scale Cessma Citation fuselage using cylindrical NAH. The tests were conducted in the Laboratory of Structural Acoustics at the Naval Research Laboratory, Washington, DC. Measurements were taken using a scanning robot to sample input data in space at a very fine step. The robot was instrumented with laser Doppler vibrometers and microphones to measure the normal surface velocity and the acoustic pressure, respectively, through an automated scanning.

In analyzing vibroacoustic responses of a complex structure, especially inside an aircraft cabin or a vehicle passenger compartment, the structural-acoustic coupling may be an important factor because the radiated acoustic pressure can act back on a structure and affect its vibrations. To acquire a better understanding of the coupling mechanisms, Kim and Kim[130] tested a partially opened plate-cavity system that allowed for two acoustic fields: one being a cavity type in which the acoustic field was reverberant, and the other being a semi-infinite type in which sound waves were propagating. Results indicated that depending on the frequency and associated wavelength, there were two types of coupling: one being a strong coupling with strong acoustic radiation and the other being a weak coupling in which the radiation efficiency was very low.

Needless to say, NAH can be used to visualize sounds emitted by a music instrument. Wang and Burroughs [131] showed examples of visualizing sound radiation from a bowed violin from 294 to 3000 Hz. The violins that were tested included the common student violin by Scherl and Roth, Hutchins violin SUS295, and Hutchins mezzo violin. The violins were bowed continuously with an open-frame mechanical bowing machine while acoustic pressure measurements were taken on different planes. The reconstructed acoustic intensity map clearly indicated the specific locations from which violin sound was radiated to the surrounding medium with spatial resolution smaller than the acoustic wavelength.

There are other recent developments in reconstruction of an acoustic field. One deals with the statistically optimized near-field acoustical holography,[132,133] which uses the same principle as that of LSM and regularization to overcome some difficulties encountered in Fourier transform based NAH. Another employs a wavelet preprocessing of the acoustic pressure signals to lessen the impact of truncation of a hologram.[134] This wavelet preprocessing employs a multiresolution analysis of sound fields for different wave number bands, followed by selective spatial filtering. Results demonstrate that this approach can reduce distortions in reconstruction due to a truncation in hologram measurements.

Other interesting developments include studies in which an acoustic field is reconstructed by taking both acoustic

pressure and in-plane particle velocity measurements.[135] The objective is to examine the feasibility of using both the particle velocity and acoustic pressure as input data in reconstruction. A pressure-velocity (P-U) probe was used to measure simultaneously the acoustic pressures and two orthogonal in-plane particle velocities.

In a separate study, comparisons of the accuracy and efficiency in reconstructing acoustic pressures and normal surface velocities were made between conventional NAH using an array of microphones and using a P-U probe.[136] Results show that when the acoustic pressure and particle velocity information are available, measurement points can be reduced by roughly 70%, which corresponds to 12% reduction in channel counts because each P-U probe contains three channels.

Williams *et al.*[137] demonstrated a measurement tool that consisted of a mobile array of 50 microphones in the form a sphere of a finite radius to visualize the flow of the acoustic intensity. By measuring the pressure signals at all microphones simultaneously and using the orthogonality condition of the spherical harmonic functions, a volumetric vector intensity field can be obtained and the results can be converted to the time domain and displayed in real time.

Bai and Lin[138] presented a near-field equivalence source imaging (NESI) method to reveal the locations and strengths of acoustic sources. This NESI is derived from time-domain formulations suitable for not only stationary but also transient acoustic fields. Multichannel inverse filters are designed and a least-square optimization is used in this approach.

Valdivia *et al.*[139] proposed two approximate methods that enabled one to carry out patch reconstruction for an arbitrary surface. The first approach is called patch IBEM with continuation procedure, and the second is known as patch IBEM with direct procedure. These methods can be very helpful in practice because sometimes measurements are available over a finite surface area only. Validations of these approximate methods are shown through numerical simulations over a cylindrical surface with the input data generated by point sources, and experimental data from a cylindrical fuselage excited by a point force.

Yu *et al.*[140] developed an acoustic intensity-based method (AIBM) to reconstruct acoustic pressures in a free field. Unlike other methods, AIBM utilizes the acoustic pressure gradient and acoustic pressure as input data. They demonstrated that the reconstructed acoustic pressures based on a combination of the acoustic pressure gradient and acoustic pressure may be more stable and less dependable on the locations of the input data than that based on the acoustic pressures as the input data alone due to the unique continuation theory of elliptic equation. Furthermore, AIBM is applicable for both forward and inverse problems up to a minimum sphere that circumscribes the acoustic sources of interest.

XI. REGULARIZATION TECHNIQUES

All inverse problems are mathematically ill posed, namely, we cannot uniquely determine the source characteristics based on measured data alone. This ill-posedness difficulty is reflected in the condition number of the transfer matrix or in the ratio of the maximum to minimum singular values of this transfer matrix. The larger the condition number, the more ill conditioned a matrix. For an ill-conditioned transfer matrix, a small error in the input data can yield disproportionately large errors in reconstruction.[141]

Solution strategies to inverse problems can be summarized in one word: regularization.[142] The underlying principle in regularization is to smooth the dependence of the output data on the input data.[143–145] For a review of various regularization techniques in NAH, readers are referred to the paper by Williams.[146] Note that there is no single regularization that is absolutely superior to the rest. It all depends on the nature and physics of the problem at hand.

The reconstruction formulations discussed above, for example, Eqs. (8) and (10), may be recast in a generic matrix equation as

$$\mathbf{p}^{\delta}(\mathbf{x}_m) = \mathbf{T}_v(\mathbf{x}_m|\mathbf{x}_S)\mathbf{v}^{\delta}(\mathbf{x}_S), \tag{14}$$

where $\mathbf{p}^{\delta}(\mathbf{x}_m)$ is the acoustic pressure measured at $\mathbf{x}_m$, $\mathbf{v}^{\delta}(\mathbf{x}_S)$ is the normal surface velocity on the source surface $\mathbf{x}_S$, and $\mathbf{T}_v$ is a (Hermitian) transfer matrix.

The simplest regularization is TSVD that truncates the transfer matrix $\mathbf{T}_v$ to include only the dominant acoustic modes. TSVD is equivalent to projecting a solution to a significant subspace spanned by a finite set of singular vectors. Hence, TSVD yields a conditionally correct solution. In most practical situations, the singular values of any transfer matrix decrease monotonically as its order increases, and the truncation order is chosen by comparing the norms of noise and those of residual errors. So TSVD solutions are determined by measurement devices or, more precisely, by the width of the singular value spectrum.

An alternative is to use weighted SVD (WSVD).[147] The drawback of WSVD is that it has no simple functional interpretation and, therefore, is less popular than TSVD. A better approach is to define functionals $J_1(\mathbf{v},\mathbf{v}_0^{\delta})$ and $J_2(\mathbf{v},\mathbf{v}_{\infty}^{\delta})$ for the true image $\mathbf{v}$ between two extreme cases $\mathbf{v}_0^{\delta}$ and $\mathbf{v}_{\infty}^{\delta}$,[148,149] where $\mathbf{v}_0^{\delta}$ represents an ultrarough least-squares solution and $\mathbf{v}_{\infty}^{\delta}$ signifies an ultrasmooth estimate of $\mathbf{v}$. The goal is to strike a balance between a good match between $\mathbf{v}_0^{\delta}$ and the measured data as described by $J_1(\mathbf{v},\mathbf{v}_0^{\delta})$ and that between $\mathbf{v}_{\infty}^{\delta}$ and the prior knowledge of the source field $\mathbf{v}$ as described by $J_2(\mathbf{v},\mathbf{v}_{\infty}^{\delta})$. This goal can be achieved by minimizing the functional

$$J(\mathbf{v}^{\delta,\alpha}) = \min_{\mathbf{v}\in\mathbf{V}}[J_1(\mathbf{v},\mathbf{v}_0^{\delta}) + \alpha J_2(\mathbf{v},\mathbf{v}_{\infty}^{\delta})], \tag{15}$$

where $\mathbf{v}^{\delta,\alpha}$ depends on a regularization parameter α, and the functionals $J_1(\mathbf{v},\mathbf{v}_0^{\delta})$ and $J_2(\mathbf{v},\mathbf{v}_{\infty}^{\delta})$ define a path between two extreme cases $\mathbf{v}_0^{\delta}$ and $\mathbf{v}_{\infty}^{\delta}$ for all possible fields.[148] The choice of the regularization parameter α allows users to decide how far to go along this path to achieve certain smoothness. For example, a good match with measured data may be achieved when $\alpha=0$, while a good match with some prior knowledge of a source field may be achieved when $\alpha=\infty$.

Equation (15) represents a specific regularization format. Once this format is selected, the next step is to determine the value of α. One way is to utilize the Morozov discrepancy

principle (MDP),[150] which states that the reconstructed pressure $\mathbf{p}^{\delta,\alpha}$ must satisfy a discrepancy equation

$$\|\mathbf{p}^{\delta,\alpha}-\mathbf{p}^{\delta}\|_2=\sigma\sqrt{M}, \tag{16}$$

where $\mathbf{p}^{\delta,\alpha}$ depends on the value of α, σ implies the standard deviation of measurement noise, and M is the number of measurement points. The standard deviation σ can be estimated by measured data, so long as the noise is Gaussian and spatially uncorrelated.

If SVD is used to reconstruct an acoustic field, we can estimate the standard deviation σ by computing the norm of the filtered pressure modes over the entire k-space as follows:[146]

$$\sigma=\|UF_h(\alpha)U^H\mathbf{p}^{\delta}\|_2/\sqrt{M}, \tag{17}$$

where $F_h(\alpha)$ is an α-dependent high-pass filter for reconstructing acoustic pressure, which passes small singular values but suppresses large ones. Once the standard deviation σ is determined, we vary the regularization parameter α in Eq. (17) until the reconstructed acoustic pressure differs from the measured data by the noise level. An improved variant of the discrepancy principle is Gfrerer and Raus' method.[151,152] Like MDP, this method needs *a priori* knowledge of noise variance.

There are other methods for calculating the value of α that require no prior knowledge of noise variance such as the predicted sum of squares method,[153] a maximum likelihood method,[154] ordinary cross validation (OCV) method,[155] generalized cross validation (GCV) method,[156] and L-curve method[157] that is based on a plot of the solution norm versus the residual norm in a log-log scale.[158] Here we discuss OCV and GCV methods because they are most often used in NAH.

The basic idea of OCV is to omit a measured acoustic pressure out of calculations of the cost function and then evaluate the effectiveness of a reconstructed source field in predicting the value of the omitted data. This OCV process is repeated for all data points omitted. The resulting regularization parameter α ensures a best fit of the predicted acoustic pressures to measured data. However, OCV may fail to yield a regularization parameter α when the matrix for predicting the missing acoustic pressure becomes diagonal. This is because when the matrix is nearly diagonal, the corresponding minimizing function reduces to the sum of squared measured acoustic pressure divided by the number of measurements, which is independent of the value of α. To circumvent this difficulty, Golub *et al.*[156] developed a GCV, which states that any good choice of α is invariant under rotation of the measurement coordinate systems. Examples of reconstructing a source field on a simply supported baffled plate using OCV and GCV together with TR are shown by Nelson and Yoon[159] and Yoon and Nelson.[160]

To summarize, we can write reconstruction formulations in the following manner:

$$\mathbf{p}^{\delta,\alpha}=\boldsymbol{V}_p\boldsymbol{F}_p^{\alpha}\boldsymbol{\Sigma}_p^{-1}\boldsymbol{U}_p^H\mathbf{p}^{\delta} \quad \text{and} \quad \mathbf{v}^{\delta,\alpha}=\boldsymbol{V}_v\boldsymbol{F}_v^{\alpha}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{U}_v^H\mathbf{p}^{\delta}, \tag{18}$$

where the reconstructed acoustic pressure $\mathbf{p}^{\delta,\alpha}$ and normal velocity $\mathbf{v}^{\delta,\alpha}$ on a source surface are α dependent, the subscripts p and v indicate reconstructions for $\mathbf{p}^{\delta,\alpha}$ and $\mathbf{v}^{\delta,\alpha}$, respectively, $\boldsymbol{\Sigma}_p^{-1}$ and $\boldsymbol{\Sigma}_v^{-1}$ represent inversions of the diagonal matrices containing singular values of transfer matrices $\mathbf{T}_p$ and $\mathbf{T}_v$ defined in Eqs. (9a) and (9b), respectively, and $\boldsymbol{F}_p^{\alpha}$ and $\boldsymbol{F}_v^{\alpha}$ imply the low-pass filters to eliminate high wave numbers (evanescent waves) in reconstructing $\mathbf{p}^{\delta,\alpha}$ and $\mathbf{v}^{\delta,\alpha}$, respectively. Note that since the singular values in $\boldsymbol{\Sigma}_p^{-1}$ and $\boldsymbol{\Sigma}_v^{-1}$ are different, the regularization parameters are different. As a result, the low-pass filters $\boldsymbol{F}_p^{\alpha}$ and $\boldsymbol{F}_v^{\alpha}$ will be different.

TABLE II. Regularization techniques for reconstructing $\mathbf{p}^{\delta,\alpha}$ and $\mathbf{v}_n^{\delta,\alpha}$.

Regularization	Filters F_p^{α} and F_v^{α}
Standard TR	$F_{p,S}^{\alpha}=\mathrm{diag}\left(\cdots, \dfrac{\varepsilon_{pj}^2}{\alpha^2+\varepsilon_{pj}^2}, \cdots\right)$
	$F_{v,S}^{\alpha}=\mathrm{diag}\left(\cdots, \dfrac{\varepsilon_{vj}^2}{\alpha^2+\varepsilon_{vj}^2}, \cdots\right)$
Modified TR	$F_{p,M}^{\alpha}=\mathrm{diag}\left(\cdots, \dfrac{\varepsilon_{pj}^2}{\alpha^3/(\alpha+\varepsilon_{pj}^2)^2+\varepsilon_{pj}^2}, \cdots\right)$
	$F_{v,M}^{\alpha}=\mathrm{diag}\left(\cdots, \dfrac{\varepsilon_{vj}^2}{\alpha^3/(\alpha+\varepsilon_{vj}^2)^2+\varepsilon_{vj}^2}, \cdots\right)$
Landweber iterations	$F_{p,L}^{\alpha}=\mathrm{diag}(\cdots, 1-(1-\alpha\varepsilon_{pj}^2)^m, \cdots)$
	$F_{v,L}^{\alpha}=\mathrm{diag}(\cdots, 1-(1-\alpha\varepsilon_{vj}^2)^m, \cdots)$

Table II lists the low-pass filters that are associated with a standard TR,[148] modified TR, and Landweber iterations.[161] The symbols ε_{pj} and ε_{vj} in Table II indicate the jth singular values of $\mathbf{T}_p$ and $\mathbf{T}_v$, respectively. The second subscripts S, M, and L in filters $F_{p,S}^{\alpha}$, $F_{p,M}^{\alpha}$, and $F_{p,L}^{\alpha}$ indicate a standard TR, modified TR, and Landweber iteration, respectively.

The main difference between a standard TR and modified TR is the penalty function. The former tends to limit the growth of all wave numbers; whereas the latter only restricts the growth of evanescent waves, but leaves low wave numbers unchanged. Since it is the evanescent waves that blow up reconstruction, a modified TR seems more logical than a standard TR.

In summary, all regularization techniques share a common goal of defining the shape of a low-pass filter to eliminate the evanescent waves and to smooth the dependence of reconstructed acoustic quantities on the input data. Each filter shape has a unique slope at the inflection point, which is defined at the point where the filter value is exactly one-half. The filter shape, especially the slope at the inflection point, determines the reconstruction accuracy. The steeper the slope is, the more accurate the reconstruction becomes. For example, standard TR has the smallest slope among all regularization listed in Table II. Accordingly, its accuracy is the worst in comparison with that of a modified TR and Landweber iterations. However, one should not conclude that a rectangular filter (for example, TSVD) whose slope at the inflection point is infinite would yield the most accurate result. On the contrary, a rectangular filter gives an even worse reconstruction than standard TR does, even though it is very convenient to use in practice. There seems to be an optimal filter shape. Among all the filters tested so far, the best is an exponential filter[146] that has a steeper slope than all the filters given in Table II and offers better reconstruction than all

TABLE III. Minimizations used in conjunction with different regularizations for determining α.

Minimization	Standard TR	Modified TR	Landweber
MDP	$\|F_{h,S}^{\alpha}U^H\mathbf{p}^{\delta}\|_2=\sigma\sqrt{M}$	$\|F_{h,M}^{\alpha}U^H\mathbf{p}^{\delta}\|_2=\sigma\sqrt{M}$	$\|F_{h,L}^{\alpha}U^H\mathbf{p}^{\delta}\|_2=\sigma\sqrt{M}$
OCV	$\min\left(\frac{\|B_S^{\alpha}F_{h,S}^{\alpha}U^H\mathbf{p}^{\delta}\|_2^2}{M}\right)$	$\min\left(\frac{\|B_M^{\alpha}F_{h,M}^{\alpha}U^H\mathbf{p}^{\delta}\|_2^2}{M}\right)$	$\min\left(\frac{\|B_L^{\alpha}F_{h,L}^{\alpha}U^H\mathbf{p}^{\delta}\|_2^2}{M}\right)$
GCV	$\min\left(\frac{\|F_{h,S}^{\alpha}U^H\mathbf{p}^{\delta}\|_2^2}{[\mathrm{Tr}(F_{h,S}^{\alpha})]^2}\right)$	$\min\left(\frac{\|F_{h,M}^{\alpha}U^H\mathbf{p}^{\delta}\|_2^2}{[\mathrm{Tr}(F_{h,M}^{\alpha})]^2}\right)$	$\min\left(\frac{\|F_{h,L}^{\alpha}U^H\mathbf{p}^{\delta}\|_2^2}{[\mathrm{Tr}(F_{h,L}^{\alpha})]^2}\right)$

other filters do. The shortcoming of this exponential filter, however, is that the parameters involved in this filter cannot be determined automatically, but rather through trial and error.

Specifying the shape of a low-pass filter only accomplishes half of the regularization. The other half is to determine the break point at which the wave numbers are truncated to avoid noise contamination. This can be done automatically by MDP when a noise variance is obtained or by GCV when no prior knowledge is available. The break point provided by MDP and GCV may be different. The former tends to give a value smaller than the critical wave number, while the latter yields a value larger than the critical one. A small break point means an oversmoothing that may discard useful information and yield a reconstructed image of low spatial resolution. On the other hand, a large break point represents an under smoothing that may allow noise contaminated high wave numbers to sneak into the reconstruction process and lead to a distorted image.

Table III provides combinations of various regularization techniques for determining the regularization parameter α. It is interesting to note that GCV may fail to produce an α when it is coupled with a standard TR. On the other hand, MDP is always capable of producing an α, but the resulting accuracy in reconstruction may not be very good when it is coupled with a standard TR. A possible compromise seems to be combining a modified TR with either GCV or MDP, which can produce a value of α and satisfactory reconstruction.[91,92,162,163] It is important to understand that there is no single regularization technique that can offer satisfactory results for all cases.

Another regularization used in NAH is the conjugate gradient least squares[164] (CGLS) or conjugate gradient for the normal equation (CGNE) method. The reason that Table II did not include these methods is because they cannot be expressed mathematically in a concise form. CGLS consists of five iterative steps that produce least-squares solution for a source field at each step. In general, CGLS can produce an optimal solution within a few iterations, but then diverges rapidly. This is because CGLS is highly sensitive to noise in the input data. Initially, the iteration solution is confined to low wave numbers that are basically noise-free. As iterations continue, the noise contaminated high wave numbers may get involved, causing solution to diverge. This rapid converging/diverging behavior makes the choices of a regularization parameter and stopping rule extremely critical to the success of CGLS. CGNE has been used to reconstruct the acoustic field through a single layer potential formulation in two dimensions,[165] and the acoustic field in 3D space by solving a combined single and double layer potential formulations and the Helmholtz integral formulations.[165]

The main advantage of CGNE is the saving in computation time, which is of order $O(N^2)$ per each iteration step in calculating a residual norm as compared to the computation time using TSVD with an order $O(N^3)$. The regularization parameter in CGNE is the number of iterations l, while that in TSVD is the rank of matrix N. Since in many cases $l \ll N$, CGNE is much faster and more efficient than TSVD. The methods developed to select a stoppage l for CGNE without a prior knowledge of the variance of measurement noise are GCV,[156] Hanke and Raus' method,[166] and DeLillo and Hrycak's method.[167]

There is another methodology known as a stochastic inversion or Gaussian inversion for handling an ill-posed problem.[168] Bonnet[169] utilized a Gaussian inversion to improve the stability of inversion and compared the reconstructed normal surface velocities for a finite cylinder with those obtained by using least squares and SVD, respectively.

XII. FUTURE PROSPECTS

Acoustic imaging methods have evolved significantly over the past three decades from a simple concept of visualizing acoustic radiation to a diagnostic tool. In particular, NAH has been recognized by the U.S. Naval Research Laboratory[170] as one of the most innovative technologies in its 75 year history. Based on the pace of evolution in the past, it will be hard to predict what it will be in the next three decades. Indeed, with able researchers working diligently in this field, nothing is impossible. However, there are a few aspects that need special attention.

(1) High frequency: Most numerical methods are effective in low-to-mid frequencies, but not in the high-frequency regime. In practice, many noise problems such as break squeals are up to 10 kHz. How to handle this high-frequency case remains one of the most challenging tasks.
(2) Transient cases: This is another major challenge. Compared to time-harmonic cases, transient case is in its infancy. The main difficulty is a lack of an effective methodology to calculate the transient acoustic radiation from a structure subject to an arbitrary time-dependent excita-

tion. On the other hand, most noise and vibration problems are produced by transient excitations.

(3) Moving source: Investigation on visualizing acoustic radiation from a moving source has just started. More effective methodologies need to be developed to correct the distortions in the reconstructed spectra and acoustic fields induced by the source convectional motion.

(4) Multiple sources: This area has received an increasing attention, but more research is needed to develop an acoustic imaging technique that can not only visualize the acoustic field from a plurality of *unknown* noise sources but also identify their individual contributions.

(5) "See-through:" This perhaps is the most challenging task of all. Conventional wisdom tells us that one cannot extend reconstruction of an acoustic quantity beyond a solid surface, because mathematically the acoustic field is undefined. In practice, we want to "see" what is going on behind a solid surface. For example, in analyzing vehicle noise what we have is the measured acoustic pressure signals outside, but we want to know the root cause of the vehicle noise and how this noise gets transmitted to the exterior region.

(6) Viscous and nonlinear fluid medium: Current technologies are for an ideal fluid medium. An extension of the acoustic imaging technologies to a viscous and nonlinear fluid medium may land their application to medical diagnosis to visualize the condition of a heart or fetus inside a human body without the need of sending potentially harmful signals from the outside such as an x ray or ultrasound.

ACKNOWLEDGMENT

This work was supported by NSF Grant No. ITR/ACS-0081270.

APPENDIX: FORMULATIONS FOR MATRICES DEFINED IN EQUATIONS (9A) AND (9B)

The elements of matrices $\mathbf{M}_S(\mathbf{x}_S)$, $\mathbf{D}_S(\mathbf{x}_S)$, $\mathbf{M}(\mathbf{x})$, and $\mathbf{D}(\mathbf{x})$ in Eqs. (9a) and (9b) are expressible in terms of surface integrations over discrete elements on a source surface,

$$M_{S_{\mu\nu}}(\mathbf{x}_{S_{\mu\nu}}) = i\omega\rho_0 \int\int_{S_{\mu\nu}} \frac{e^{ikR_S}}{R_S} dS_{\mu\nu}, \tag{A1}$$

$$D_{S_{\mu\nu}}(\mathbf{x}_S) = \int\int_{S_{\mu\nu}} \left(ik - \frac{1}{R_S}\right) \frac{e^{ikR_S}}{R_S} \frac{\partial R_S}{\partial n} dS_{\mu\nu}, \tag{A2}$$

$$M_{\mu\nu}(\mathbf{x}) = i\omega\rho_0 \int\int_{S_{\mu\nu}} \frac{e^{ikR}}{R} dS_{\mu\nu}, \tag{A3}$$

$$D_{\mu\nu}(\mathbf{x}) = \int\int_{S_{\mu\nu}} \left(ik - \frac{1}{R}\right) \frac{e^{ikR}}{R} \frac{\partial R}{\partial n} dS_{\mu\nu}, \tag{A4}$$

where the subscript $S_{\mu\nu}$ indicates the $\mu\nu$th element on a source surface S, and indices μ, $\nu = 1, 2, \ldots, N$, here N is the total number of elements discretized on S.

[1]B. P. Hildebrand and B. B. Brenden, *An Introduction to Acoustical Holography* (Plenum, New York, 1974).
[2]J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).
[3]H. M. Smith, *Principles of Holography* (Wiley-Interscience, New York, 1969).
[4]E. G. Williams and J. D. Maynard, "Numerical evaluation of the Rayleigh integral for planar radiators using the FFT," J. Acoust. Soc. Am. **72**, 2020–2030 (1982).
[5]E. G. Williams and J. D. Maynard, "Holographic imaging without the wavelength resolution limit," Phys. Rev. Lett. **45**, 554–557 (1980).
[6]E. G. Williams, J. D. Maynard, and E. Skudrzky, "Sound source reconstructions using a microphone array," J. Acoust. Soc. Am. **68**, 340–344 (1980).
[7]E. G. Williams, H. D. Dardy, and R. G. Fink, "Nearfield acoustical holography using an underwater automated scanner," J. Acoust. Soc. Am. **78**, 789–798 (1985).
[8]P. R. Stepanishen and K. C. Benjamin, "Forward and backward projection of acoustic fields using FFT methods," J. Acoust. Soc. Am. **71**, 803–812 (1982).
[9]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).
[10]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, San Diego, CA, 1999).
[11]M. C. Junger and D. Feit, *Sound, Structures, and Their Interactions* (MIT, Cambridge, MA, 1972), pp. 403–432.
[12]S. Hayek and T. Luce, "Aperture effects in planar nearfield acoustical imaging," Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **110**, 91–96 (1988).
[13]E. G. Williams, B. H. Houston, and P. C. Herdic, "Fast Fourier transform and singular value decomposition formulations for patch nearfield acoustical holography," J. Acoust. Soc. Am. **114**, 1322–1333 (2003).
[14]A. Sarkissian, "Extension of measurement surface in near-field acoustic holography," J. Acoust. Soc. Am. **115**, 1593–1596 (2004).
[15]A. Sarkissian, "Method of superposition applied to patch near-field," J. Acoust. Soc. Am. **118**, 671–677 (2005).
[16]M. Lee and J. S. Bolton, "Patch near-field acoustical holography in cylindrical geometry," J. Acoust. Soc. Am. **118**, 3721–3732 (2005).
[17]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).
[18]F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," Proc. IEEE **66**, 51–83 (1978).
[19]J. W. Tukey, "An introduction to the calculations of numerical spectrum analysis," in *Spectral Analysis of Time Series*, edited by B. Harris (Wiley, New York, 1967), pp. 25–46.
[20]E. G. Williams, H. D. Dardy, and R. G. Fink, "A technique for measurement of structure-borne intensity in plates," J. Acoust. Soc. Am. **78**, 2061–2066 (1985).
[21]E. G. Williams, "Supersonic acoustic intensity," J. Acoust. Soc. Am. **97**, 121–127 (1995).
[22]T. Loyau and J.-C. Pascal, "Broadband acoustic holography reconstruction from acoustic intensity measurements. I: Principle of the method," J. Acoust. Soc. Am. **84**, 1744–1750 (1988).
[23]E. G. Williams, B. H. Houston, and J. A. Bucaro, "Broadband nearfield acoustical holography for vibrating cylinders," J. Acoust. Soc. Am. **86**, 674–679 (1989).
[24]E. G. Williams, H. D. Dardy, and K. B. Washburn, "Generalized nearfield acoustical holography for cylindrical geometry: Theory and experiment," J. Acoust. Soc. Am. **81**, 389–407 (1987).
[25]E. G. Williams, "Imaging the sources on a cylindrical shell from far field pressure measured on a semicircle," J. Acoust. Soc. Am. **99**, 2022–2032 (1996).
[26]E. G. Williams, "The nearfield acoustic holography (NAH) experimental method applied to vibration and radiation in light and heavy fluids," Comput. Struct. **65**, 323–335 (1997).
[27]W. A. Veronesi and J. D. Maynard, "Nearfield acoustic holography (NAH): II. Holographic reconstruction algorithms and computer implementation," J. Acoust. Soc. Am. **81**, 1307–1322 (1987).
[28]Y.-C. Chao, "An implicit least-squares method for the inverse problem of acoustic radiation," J. Acoust. Soc. Am. **81**, 1288–1292 (1987).
[29]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Chap. 4, pp. 165–194.
[30]A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, "Application of

the BIE method to sound radiation problems using an isoparametric element," Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **106**, 414–420 (1984).
[31]A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, "An advanced computational method for radiation and scattering of acoustic waves in three dimensions," J. Acoust. Soc. Am. **77**, 362–368 (1985).
[32]M. A. Latcha and A. Akay, "Application of the Helmholtz integral in acoustics," Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **108**, 447–453 (1986).
[33]T. W. Wu, "A direct boundary element method for acoustic radiation and scattering from mixed regular and thin bodies," J. Acoust. Soc. Am. **97**, 84–91 (1995).
[34]B. K. Gardner and R. J. Bernhard, "A noise source identification technique using an inverse Helmholtz integral equation method," ASME J. Vib., Acoust., Stress, Reliab. Des. **110**, 84–90 (1988).
[35]W. A. Veronesi and J. D. Maynard, "Digital holographic reconstruction of sources with arbitrarily shaped surfaces," J. Acoust. Soc. Am. **85**, 588–598 (1989).
[36]G. V. Borgiotti, A. Sarkissian, E. G. Williams, and L. Schuetz, "Conformal generalized near-field acoustic holography for axisymmetric geometries," J. Acoust. Soc. Am. **88**, 199–209 (1990).
[37]J. M. Varah, "On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **10**, 257–267 (1973).
[38]G.-T. Kim and B.-T. Lee, "3-D sound source reconstruction and field reproduction using the Helmholtz integral equation," J. Sound Vib. **136**, 245–261 (1990).
[39]M. R. Bai, "Application of BEM (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrarily shaped geometries," J. Acoust. Soc. Am. **92**, 533–549 (1992).
[40]B.-K. Kim and J.-G. Ih, "On the reconstruction of vibro-acoustic field over the surface enclosing an interior space using the boundary element method," J. Acoust. Soc. Am. **100**, 3030–3016 (1996).
[41]Y.-K. Kim and Y.-H. Kim, "Holographic reconstruction of active sources and surface admittance in an enclosure," J. Acoust. Soc. Am. **105**, 2377–2383 (1999).
[42]Z. Zhang, N. Vlahopoulos, S. T. Raveendra, T. Allen, and K. Y. Zhang, "A computational acoustic field reconstruction process based on an indirect boundary element formulation," J. Acoust. Soc. Am. **108**, 2167–2178 (2000).
[43]S.-C. Kang and J.-G. Ih, "Use of non-singular boundary integral formulation for reducing errors due to near-field measurements in the boundary element method based near-field acoustic holography," J. Acoust. Soc. Am. **109**, 1320–1328 (2001).
[44]A. Schuhmacher, J. Hald, K. B. Rasmussen, and P. C. Hansen, "Sound source reconstruction using boundary element calculations," J. Acoust. Soc. Am. **113**, 114–127 (2004).
[45]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).
[46]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**, 1559–1573 (2004).
[47]N. Valdivia and E. G. Williams, "Krylov subspace iterative methods for boundary element method based near-field acoustic holography," J. Acoust. Soc. Am. **117**, 711–724 (2005).
[48]C. Langrenne, M. Melon, and A. Garcia, "Boundary element method for the acoustic characterization of a machine in bounded noisy environment," J. Acoust. Soc. Am. **121**, 2750–2757 (2007).
[49]G. H. Golub and C. F. Loan, *Matrix Computations* (Johns Hopkins University Press, North Oxford, 1983).
[50]D. M. Photiadis, "The relationship of singular value decomposition to wave-vector filtering in sound radiation problems," J. Acoust. Soc. Am. **88**, 1152–1159 (1990).
[51]A. Sarkissian, "Near-field acoustic holography for an axisymmetric geometry: A new formulation," J. Acoust. Soc. Am. **88**, 961–966 (1990).
[52]A. Sarkissian, "Acoustic radiation from finite structures," J. Acoust. Soc. Am. **90**, 574–578 (1991).
[53]A. Sarkissian, "Reconstruction of the acoustic field on radiating structures," J. Acoust. Soc. Am. **92**, 825–830 (1992).
[54]A. N. Norris, "Far-field acoustical holography onto cylindrical surfaces using pressure measured on semicircles," J. Acoust. Soc. Am. **102**, 2098–2107 (1999).
[55]M. Villot, G. Chavériat, and J. Roland, "Phonoscopy: An acoustical holography technique for plane structures radiating in enclosed spaces," J. Acoust. Soc. Am. **91**, 187–195 (1992).
[56]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).
[57]S. F. Wu, "On reconstruction of acoustic pressure fields using the Helmholtz equation least squares method," J. Acoust. Soc. Am. **107**, 2511–2522 (2000).
[58]P. J. Davis and P. Rabinowitz, "Advances in Orthonormalizing Computation," Advances in Computers **2**, 55–133 (1961).
[59]L. Rayleigh, "On the dynamical theory of gratings," Proc. R. Soc. London, Ser. A **79**, 399–416 (1907).
[60]L. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. **2**, pp. 89–96.
[61]P. M. van den Berg and J. T. Fokkema, "The Rayleigh hypothesis in the theory of diffraction by a cylindrical obstacle," IEEE Trans. Antennas Propag. **AP-27**, 577–583 (1979).
[62]R. F. Millar, "Rayleigh hypothesis in scattering problems," Electron. Lett. **5**, 416–417 (1969).
[63]R. F. Millar, "On the Rayleigh assumption in scattering by a periodic surface," Proc. Cambridge Philos. Soc. **65**, 773–791 (1969).
[64]R. F. Millar, "The Rayleigh hypothesis and a related least-squares solution to scattering problems for periodic surfaces and other scatters," Radio Sci. **8**, 785–796 (1973).
[65]R. F. Millar, "Singularities and the Rayleigh hypothesis for solutions to the Helmholtz equation," IMA J. Appl. Math. **37**, 155–171 (1986).
[66]P. J. Davis, "The Schwarz Function and Its Applications," Math. Gaz. **59**, 127–128 (1975).
[67]H. S. Shapiro, *The Schwarz Function and Its Generalization To Higher Dimensions* (John Wiley and Sons, New York, 1992).
[68]T. Semenova and S. F. Wu, "The Helmholtz equation least squares method and the Rayleigh's hypothesis in nearfield acoustical holography," J. Acoust. Soc. Am. **115**, 1632–1640 (2004).
[69]N. E. Rayess, "Development of acoustic holography using the Helmholtz equation-least squares (HELS) method," Ph.D. thesis, Department of Mechanical Engineering, Wayne State University, pp. 65–77 (2001).
[70]V. Isakov and S. F. Wu, "On theory and applications of the HELS method in inverse acoustics," Inverse Probl. **18**, 1147—1159 (2002).
[71]A. N. Tikhonov, "On stability of inverse problems," Dokl. Akad. Nauk SSSR **39**, 176–179 (1943).
[72]T. Semenova, "On the behavior of HELS solutions for acoustic radiation and reconstruction," Ph.D. thesis, Department of Mechanical Engineering, Wayne State University, pp. 113–123 (2004).
[73]J. Prager, "Approximate reconstruction of sound fields close to the source surface using spherical nearfield acoustic holography," J. Acoust. Soc. Am. **122**, 2067–2073 (2007).
[74]N. E. Rayess and S. F. Wu, "Experimental validations of the HELS method for reconstructing acoustic radiation from a complex vibrating structure," J. Acoust. Soc. Am. **107**, 2955–2964 (2000).
[75]S. F. Wu, N. Rayess, and X. Zhao, "Visualization of acoustic radiation from a vibrating bowling ball," J. Acoust. Soc. Am. **109**, 2771–2779 (2001).
[76]S. F. Wu and J. Yu, "Reconstructing interior acoustic pressure field via Helmholtz equation least-squares method," J. Acoust. Soc. Am. **104**, 2054–2060 (1998).
[77]M. Moondra and S. F. Wu, "Visualization of vehicle interior sound field using HELS based NAH," Noise Control Eng. J. **53**, 146–154 (2005).
[78]D. C. Kammer, "Sensor placement for on-orbit modal identification and correlation of large space structures," J. Guid. Control Dyn. **14**, 251–259 (1991).
[79]D. C. Kammer, "Effect of modal error on sensor placement for on-orbit modal identification of large space structures," J. Guid. Control Dyn. **15**, 334–341 (1992).
[80]D. C. Kammer, "Effects of noise on sensor placement for on-orbit modal identification of large space structures," ASME J. Dyn. Syst., Meas., Control **114**, 436–443 (1992).
[81]B.-K. Kim and J.-G. Ih, "On the reconstruction of the vibro-acoustic field over the surface enclosing in interior space using the boundary element method," J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[82]B.-K. Kim and J.-G. Ih, "Design of an optimal wave-vector filter for enhancing the resolution of reconstructed source field by near-field acoustical holography (NAH)," J. Acoust. Soc. Am. **107**, 3289–3297 (2000).
[83]G. P. Carroll, "The effect of sensor placement errors on cylindrical near-field acoustic holography," J. Acoust. Soc. Am. **105**, 2269–2276 (1999).

[84] K.-U. Nam and Y.-H. Kim, "Errors due to sensor and position mismatch in planar acoustic holography," J. Acoust. Soc. Am. **106**, 1655–1665 (1999).
[85] H. C. Andrews and B. R. Hunt, *Digital Image Restoration* (Prentice-Hall, Englewood Cliffs, NJ, 1977).
[86] H. Fleischer and V. Axelrad, "Restoring an acoustic source from pressure data using Wiener filtering," Acustica **60**, 172–175 (1986).
[87] M. R. Bai, "Acoustical source characterization by using recursive Wiener filtering," J. Acoust. Soc. Am. **97**, 2657–2663 (1995).
[88] S.-C. Kang and J.-G. Ih, "The use of partially measured source data in near-field acoustical holography based on the BEM," J. Acoust. Soc. Am. **107**, 2472–2479 (2000).
[89] S. F. Wu and X. Zhao, "Combined Helmholtz equation least squares (CHELS) method for reconstructing acoustic radiation," J. Acoust. Soc. Am. **112**, 179–188 (2002).
[90] S. F. Wu, "Hybrid nearfield acoustical holography," J. Acoust. Soc. Am. **115**, 207–217 (2004).
[91] X. Zhao and S. F. Wu, "Reconstruction of vibro-acoustic fields by using hybrid nearfield acoustical holography in half space," J. Acoust. Soc. Am. **117**, 555–565 (2005).
[92] X. Zhao and S. F. Wu, "Reconstruction of the vibro-acoustic fields using hybrid nearfield acoustical holography," J. Sound Vib. **282**, 1183–1199 (2005).
[93] J. A. Mann III, E. G. Williams, and K. Washburn, "Time-domain analysis of the energy exchange between structural vibrations and acoustic radiation using near-field acoustical holography measurements," J. Acoust. Soc. Am. **90**, 1656–1664 (1991).
[94] T. B. Hansen, "Spherical expansions of time-domain acoustic fields: Application to near-field scanning," J. Acoust. Soc. Am. **98**, 1204–1215 (1995).
[95] O. d. L. Rochefoucauld, M. Melon, and A. Garcia, "Time domain holography: Forward projection of simulated and measured sound pressure fields," J. Acoust. Soc. Am. **116**, 142–151 (2004).
[96] S. F. Wu, H. Lu, and M. S. Bajwa, "Reconstruction of transient acoustic radiation from a sphere," J. Acoust. Soc. Am. **117**, 2065–2077 (2005).
[97] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986), pp. 332–347.
[98] S. F. Wu, "Transient radiation from an impulsively accelerated body," J. Acoust. Soc. Am. **94**, 542–554 (1993).
[99] M. Ochmann, "Comments on 'Transient radiation from an impulsively accelerated body,'" J. Acoust. Soc. Am. **97**, 3895–3896 (1995).
[100] S. F. Wu, "Response to 'Comments on transient radiation from an impulsively accelerated body,'" J. Acoust. Soc. Am. **97**, 3897 (1995).
[101] H.-S. Kwon and Y.-H. Kim, "Moving frame technique for planar acoustic holography," J. Acoust. Soc. Am. **103**, 1734–1741 (1998).
[102] S.-H. Park and Y.-H. Kim, "An improved moving frame acoustic holography for coherent bandlimited noise," J. Acoust. Soc. Am. **104**, 3179–3189 (1998).
[103] X.-F. Wu and A. Akay, "Sound radiation from vibrating bodies in motion," J. Acoust. Soc. Am. **91**, 2544–2555 (1992).
[104] S.-H. Park and Y.-H. Kim, "Effects of the speed of moving noise sources on the sound visualization by means of moving frame acoustic holography," J. Acoust. Soc. Am. **108**, 2719–2728 (2000).
[105] S.-H. Park and Y.-H. Kim, "Visualization of pass-by noise by means of moving frame acoustic holography," J. Acoust. Soc. Am. **110**, 2326–2339 (2001).
[106] R. J. Ruhala and D. C. Swanson, "Planar near-field acoustical holography in a moving medium," J. Acoust. Soc. Am. **112**, 420–429 (2002).
[107] M. Ochmann, "The source simulation technique for acoustic radiation problems," Acustica **81**, 512–527 (1995).
[108] I.-Y. Jeon and J.-G. Ih, "On the holographic reconstruction of vibroacoustic fields using equivalent sources and inverse boundary element method," J. Acoust. Soc. Am. **118**, 3475–3484 (2005).
[109] G. H. Koopmann, L. Song, and J. B. Fahnline, "A method for computing acoustic fields based on the principle of wave superposition," J. Acoust. Soc. Am. **86**, 2433–2438 (1989).
[110] L. Song, G. H. Koopmann, and J. B. Fahnline, "Numerical errors associated with the method of superposition for computing acoustic fields," J. Acoust. Soc. Am. **89**, 2625–2633 (1991).
[111] J. B. Fahnline and G. H. Koopmann, "A numerical solution for the general radiation problem based on the combined methods of superposition and singular-value decomposition," J. Acoust. Soc. Am. **90**, 2808–2819 (1991).
[112] K. H. Baek and S. J. Elliot, "Natural algorithms for choosing source locations in active control systems," J. Sound Vib. **186**, 245–267 (1995).
[113] T. Semenova and S. F. Wu, "On the choice of expansion functions in the Helmholtz equation least-squares method," J. Acoust. Soc. Am. **117**, 701–710 (2005).
[114] A. Doicu, Y. Eremin, and T. Wriedt, *Acoustic and Electromagnetic Scattering Analysis Using Discrete Sources* (Academic, New York, 2000).
[115] Y. A. Eremin, "Complete systems of functions for the study of boundary value problems in mathematical physics," Sov. Phys. Dokl. **32**, 635–637 (1987).
[116] V. D. Kupradze, "On the approximate solutions of problems in mathematical physics," Russ. Math. Surveys **22**, 58–108 (1967).
[117] M. A. Tomlinson, "Partial source discrimination in near field acoustic holography," Appl. Acoust. **57**, 243–261 (1999).
[118] S. H. Yoon and P. A. Nelson, "A method for the efficient construction of acoustic pressure cross-spectral matrices," J. Sound Vib. **233**, 897–920 (2000).
[119] R. J. Alfredson, "The partial coherence technique for source identification on a diesel engine," J. Sound Vib. **55**, 487–494 (1977).
[120] M. E. Wang and M. J. Crocker, "On the application of coherence techniques for source identification in a multiple noise source environment," J. Sound Vib. **74**, 861–872 (1983).
[121] K.-U. Nam and Y.-H. Kim, "Visualization of multiple incoherent sources by the backward prediction of near-field acoustic holography," J. Acoust. Soc. Am. **109**, 1808–1816 (2001).
[122] J. M. Danthez and R. Aquilina, "Separation of broadband sources processing concept of the Labrador software," Mech. Syst. Signal Process. **11**, 91–106 (1997).
[123] H.-S. Kwon, Y.-J. Kim, and J. S. Bolton, "Compensation for source nonstationary in multireference, scan-based near-field acoustical holography," J. Acoust. Soc. Am. **113**, 360–368 (2003).
[124] M. Lee, J. S. Bolton, and L. Mongeau, "Application of cylindrical near-field acoustical holography to the visualization of aeroacoustic sources," J. Acoust. Soc. Am. **114**, 842–858 (2003).
[125] M. Lee and J. S. Bolton, "Scan-based near-field acoustical holography and partial field decomposition in the presence of noise and source level variation," J. Acoust. Soc. Am. **119**, 382–393 (2006).
[126] Y.-J. Kim, J. S. Bolton, and H.-S. Kwon, "Partial sound field decomposition in multireference near-field acoustical holography by using optimally located virtual references," J. Acoust. Soc. Am. **115**, 1641–1652 (2004).
[127] K.-U. Nam and Y.-H. Kim, "A partial field decomposition algorithm and its examples for near-field acoustic holography," J. Acoust. Soc. Am. **116**, 172–185 (2004).
[128] E. G. Williams, B. H. Houston, P. C. Herdic, S. T. Raveendra, and B. Gardner, "Interior nearfield acoustical holography in flight," J. Acoust. Soc. Am. **108**, 1451–1463 (2000).
[129] P. C. Herdic, B. H. Houston, M. H. Marcus, and E. G. Williams, "The vibro-acoustic response and analysis of a full-scale aircraft fuselage section for interior noise reduction," J. Acoust. Soc. Am. **117**, 3667–3678 (2005).
[130] S.-M. Kim and Y.-H. Kim, "Structural-acoustic coupling in a partially opened plate-cavity system: Experimental observation by using nearfield acoustic holography," J. Acoust. Soc. Am. **109**, 65–74 (2001).
[131] L. M. Wang and C. B. Burroughs, "Acoustic radiation from bowed violins," J. Acoust. Soc. Am. **110**, 543–555 (2001).
[132] 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).
[133] 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).
[134] J.-H. Thomas and J.-C. Pascal, "Wavelet preprocessing for lessening truncation effects in nearfield acoustical holography," J. Acoust. Soc. Am. **118**, 851–860 (2005).
[135] F. Jacobsen and Y. Liu, "Near field acoustic holography with particle velocity transducers," J. Acoust. Soc. Am. **118**, 3139–3144 (2005).
[136] M. C. Harris, J. D. Blotter, and S. D. Sommerfeldt, "Obtaining the complex pressure field at the hologram surface for use in near-field acoustical holography when pressure and in-plane velocities are measured," J. Acoust. Soc. Am. **119**, 808–816 (2006).
[137] E. G. Williams, N. Valdivia, P. C. Herdic, and J. Klos, "Volumetric acoustic vector intensity imager," J. Acoust. Soc. Am. **120**, 1887–1896 (2006).
[138] M. R. Bai and J.-H. Lin, "Source identification system based on the time-domain nearfield equivalence source imaging: Fundamental theory and implementation," J. Sound Vib. **307**, 202–225 (2007).

[139] N. P. Valdivia, E. G. Williams, and P. C. Herdic, "Approximations of inverse boundary element methods with partial measurements of the pressure field," J. Acoust. Soc. Am. **123**, 109–120 (2008).
[140] C. Yu, Z. Zhou, and M. Zhuang, "An acoustic intensity-based method for reconstruction of radiated fields," J. Acoust. Soc. Am. **123**, 1892–1901 (2008).
[141] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," IEEE Trans. Acoust., Speech, Signal Process. **37**, 2024–2036 (1989).
[142] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, New York, 1977), Chap. 2, pp. 71–73.
[143] D. L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," J. Assoc. Comput. Mach. **9**, 84–97 (1962).
[144] S. Twomey, "On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature," J. Assoc. Comput. Mach. **10**, 97–101 (1962).
[145] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia, PA, 1998).
[146] E. G. Williams, "Regularization methods for nearfield acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[147] G. Wahba, "Constrained regularization for ill-posed linear operator equations, with applications in meteorology and medicine," in *Statistical Decision Theory and Related Topics III*, edited by S. S. Gupta and J. O. Berger (Academic, New York, 1982), pp. 383–417.
[148] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," Sov. Math. Dokl. **4**, 1035–1038 (1963).
[149] D. M. Titterington, "General structure of regularization procedures in image restoration," Astron. Astrophys. **144**, 381–387 (1985).
[150] V. A. Morozov, "The error principle in the solution of operational equations by the regularization method," USSR Comput. Math. Math. Phys. **8**, 63–87 (1968).
[151] H. Gfrerer, "An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates," Math. Comput. **49**, 507–522 (1987).
[152] M. Hanke and P. C. Hansen, "Regularization method for large-scale problems," Surv. Math. Ind. **3**, 253–315 (1993).
[153] M. Allen, "The relationship between variable selection and data augmentation and a method for prediction," Technometrics **16**, 125–127 (1974).
[154] B. Anderssen and P. Bloomfield, "Numerical differentiation procedures for non-exact data," Numer. Math. **22**, 157–182 (1974).
[155] G. Wahba and S. Wold, "A completely automatic French curve: Fitting splines by cross validation," Commun. Stat. **4**, 1–17 (1975).
[156] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," Technometrics **21**, 215–223 (1979).
[157] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularisation of discrete ill-posed problems," SIAM J. Sci. Comput. (USA) **14**, 1487–1503 (1993).
[158] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," SIAM Rev. **34**, 561–580 (1992).
[159] P. A. Nelson and S. H. Yoon, "Estimation of acoustic source strength by inverse methods: Part I, Conditioning of the inverse problem," J. Sound Vib. **233**, 643–668 (2000).
[160] S. H. Yoon and P. A. Nelson, "Estimation of acoustic source strength by inverse methods: Part II. Experimental investigation of methods for choosing regularization parameters," J. Sound Vib. **233**, 669–705 (2000).
[161] L. Landweber, "An iteration formula for Fredholm integral equations of the first kind," Astronomie **73**, 615–624 (1951).
[162] K. Saijyou, "Regularization method for the application of K-space data extrapolation to near-field acoustical holography," J. Acoust. Soc. Am. **116**, 396–404 (2004).
[163] K. Saijyou and C. Okawara, "Regularization method for measurement of structural intensity using nearfield acoustical holography," J. Acoust. Soc. Am. **117**, 2039–2046 (2005).
[164] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems* (Springer-Verlag, New York, 1996).
[165] T. K. DeLillo, V. Isakov, N. Valdivia, and L. Wang, "The detection of the source of acoustical noise in two dimensions," SIAM J. Appl. Math. **61**, 2104–2121 (2001).
[166] M. Hanke, "Conjugate gradient methods for ill-posed problems," Pitmann Research Note in Mathematical Series (1995).
[167] T. K. DeLillo and T. Hrycak, "A stopping rule for the conjugate gradient regularization method for ill-posed problems" J. Comput. Acoust. **14**, 397–414 (2006).
[168] A. Tarantola, *Inversion Problem Theory* (Elsevier, New York, 1987).
[169] M. Bonnet, "A numerical investigation for a source inverse problem in linear acoustics," Journal d'Acoustique **4**, 307–334 (1991).
[170] Naval Research Laboratory, "Generalized nearfield acoustical holography," Report No. NRL/PU1001-99-393, March 2000.

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.

Linearly filtered estimation of the time-domain Green's function from measurements of ambient noise (L)

S. A. Albahrani,[a] M. R. Frater, and E. H. Huntington[b]
School of Information Technology and Electrical Engineering, University College, University of New South Wales, Canberra, ACT 2600 Australia

(Received 12 June 2008; revised 6 August 2008; accepted 6 August 2008)

It is possible to estimate the time-domain Green's function of a channel based on measurements of ambient noise by sensors at either end of the channel. This paper presents theoretical results for the impact of filtering on this problem. These results lead to the development of two experimental rules-of-thumb. It is shown that there exists a relationship between system bandwidth and sensor separation, which determines the resolvability of the measurements. The relationship between high-pass filtering and differentiation is discussed, contributing to the debate about whether or not differentiation is required to estimate the time-domain Green's function.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2981049]

PACS number(s): 43.60.Ac, 43.60.Vx, 43.20.Fn [EJS] Pages: 2699–2701

A new technique, sometimes called passive imaging,[1] has recently been used to reconstruct tomographically the seismic properties of the Earth's crust, with a significant reduction in the time taken to make these measurements compared to traditional imaging techniques based on seismic waves emitted from earthquakes.[2] Passive imaging is based on the observation that the time-domain Green's function of a channel can be reconstructed from measurements of ambient noise recorded by receivers at either end of the channel.[3–5]

Early work in this field explored this technique using ultrasound, where it was shown that the correlation of the ultrasonic diffuse field results in the local ultrasonic Green's function.[3,4,6] Since those pioneering experiments, the technique has found application in a variety of scenarios, such as in helioseismology,[7] and in seismology, where for instance, in Ref. 8, it was used in seismic exploration, and in underwater acoustics.[5,9–11]

Theoretical work has shown that the derivative of the cross-correlation yields the time-domain Green's function of the system, and not the cross-correlation function itself.[12] This is in contrast to common experimental practice where numerical differentiation is not used.[5,13,14] Reference 15 points out that numerical differentiation yields the velocity rather than the displacement Green's function, while Ref. 12 suggests that filtering may obviate the need for numerical differentiation.

The purpose of this paper is to present theoretical results for the impact of filtering on passive imaging systems. These results contribute to the debate on the need for differentiation as well as introducing the concept of resolvability in passive imaging systems. The theoretical framework may be used to develop rules-of-thumb to guide experimental design and analysis.

All of the mathematics presented herein apply to a system in which there are two sensors located at different points in an infinite, three-dimensional medium. Further, that medium contains an infinite number of randomly distributed noise sources with a uniform density. Each statistically independent noise source emits white, Gaussian noise, and is an isotropic radiator.

More specifically, the sensors are located at $(\pm\beta,0,0)$, the phase velocity in the medium is c, the power of the noise sources is Q^2, the integration time of the measurement is T, and the noise-source creation rate is v.

The average cross-correlation of the noise signals recorded at the two sensors is given by Eq. (14) from Ref. 12. After a change of variable, and dividing the integral into two parts, the cross-correlation function is

$$\langle U_{1,2}(\tau)\rangle = D[u(\tau_+) - u(\tau_-)], \tag{1}$$

where $D=c\pi Q^2TvA$, A represents the amplitude of the noise correlation function,[12] $u(t)$ is the Heaviside step function, and $\tau_\pm=\tau\pm 2\beta/c$. The derivative of the cross-correlation

[a]Present address: Macquarie University, NSW 2109 Australia.
[b]Author to whom correspondence should be addressed. Email: e.huntington@adfa.edu.au

function yields the causal and anticausal Green's function between these two points.[12]

Lemma 1: Let $h_c(t)$ be the impulse response of a filter placed immediately at the output of the cross-correlator. The output of that filter is the same as the output of a cross-correlator with two filters, $h_1(t)$ and $h_2(t)$, placed on its inputs when $h_c(t)$ is the cross-correlation of $h_1(t)$ and $h_2(t)$.

Proof: Consider the latter case. Denoting the inputs to the two filters as $n_1(t)$ and $n_2(t)$. The output of each filter (and, hence, each input to the cross-correlator) is the convolution between its impulse response function and the input. Convolution (cross-correlation) in the time domain is multiplication (multiplication by the complex conjugate) in the frequency domain.[16] Thus,

$$\begin{aligned}\tilde{C}_{1,2}(\omega) &= (\tilde{n}_1(\omega)\tilde{h}_1(\omega))(\tilde{n}_2(\omega)\tilde{h}_2(\omega))^{\star} \\ &= [\tilde{n}_1(\omega)\tilde{n}_2^{\star}(\omega)][\tilde{h}_1(\omega)\tilde{h}_2^{\star}(\omega)] = \tilde{U}_{1,2}(\omega)\tilde{h}_c(\omega).\end{aligned} \tag{2}$$

Circumflexes are used to denote Fourier transforms, $C_{1,2}(t)$ represents the output of the cross-correlator and $U_{1,2}(t)$ represents the cross-correlation between the unfiltered inputs to the system.

Lemma 1, Corollary: Using Eqs. (1) and (2), the cross-correlation function in the time domain becomes

$$\langle C_{1,2}(\tau)\rangle = D\left[\int_0^{\tau_+} du h_c(u) - \int_0^{\tau_-} du h_c(u)\right]. \tag{3}$$

Definition 1: The first (second) term in the derivative of Eq. (3) is defined to be the image of the anti-causal (causal) Green's function.

Definition 2: The sensor separation at which the first null of the image of the causal Green's function overlaps with the first null of the image of the anti-causal Green's function is defined to be the critical sensor separation.

Definition 3: The two peaks of the derivative of the cross-correlation function can be resolved when the sensors are separated by at least the critical sensor separation.

Lemma 2: Let the filter placed immediately at the output of the cross-correlator be a sinc-in-time low-pass filter with cut-off frequency f_0. The locations of the two peaks of the derivative of the correlation function can be resolved when $f_0 \geqslant c/(4\beta)$.

Proof: The impulse response of a sinc-in-time low-pass (or brickwall) filter is

$$h_{\mathrm{LPF}}(t) = 2f_0 \mathrm{sinc}(2f_0 t), \tag{4}$$

where $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$ is the normalized sinc function. Substitution into Eq. (3) and differentiation yields

$$\frac{d}{d\tau}\langle C_{1,2}(\tau)\rangle = 2Df_0[\mathrm{sinc}(f_0\tau_+) - \mathrm{sinc}(f_0\tau_-)]. \tag{5}$$

According to Definition 1, the first term in Eq. (5) is the image of the anti-causal Green's function and the second is the image of the causal Green's function. Each image is a sinc function with a half main-lobe width of $1/2f_0$, centered at $\tau \mp 2\beta/c$, respectively. The critical sensor separation is therefore given by $2\beta = c/2f_0$ according to Definition 2.

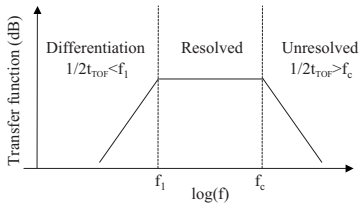


FIG. 1. The frequency characteristic of a band-pass filter and the corresponding measurement regimes.

Hence, the locations of the causal and anti-causal Green's functions can be resolved when $2\beta > c/2f_0$ or equivalently $f_0 > c/4\beta$ as per definition 3.

Lemma 3: Let the filter placed immediately at the output of the cross-correlator be a first-order high-pass filter with cut-off frequency f_1. This filtering can take the place of linear scaling followed by differentiation of the correlation function when $f_1 \gg c/(4\beta)$.

Proof: The impulse response function of the first order high pass filter is

$$h_{\mathrm{foHPF}} = \delta(t) - 2\pi f_1 e^{-2\pi f_1 t}u(t). \tag{6}$$

Substitution into Eq. (3) and using $f_1 \gg c/(4\beta)$ yields

$$\begin{aligned}\langle C_{1,2}(\tau)\rangle &= D[u(\tau_+)e^{-2\pi f_1\tau_+} - u(\tau_-)e^{-2\pi f_1\tau_-}] \\ &\approx \frac{D}{\omega_1}[\delta(\tau_+) - \delta(\tau_-)].\end{aligned} \tag{7}$$

□

Two rules-of-thumb are proposed for the application of Lemmas 2 and 3 to practical situations.

The limitation on the upper cut-off frequency of the sensor system (Lemma 2) can be expressed as: *The response time of the measurement system must be less than twice the time-of-flight between the sensors.*

The relation given in Lemma 3 can be expressed as: *Significant differentiation is contributed by the sensors when time-constant of the high-pass filter is much less than twice the time-of-flight between the sensors.*

Practical sensors tend to exhibit a band-pass characteristic as illustrated in Fig. 1. This characteristic, in combination with the proposed rules-of-thumb, leads to the three regions of time-of-flight between the sensors, $t_{\mathrm{TOF}} = 2\beta/c$. The three regions are: the differentiation region, $1/2t_{\mathrm{TOF}} < f_1$, in which the transfer function acts as a high-pass filter, and hence, contributes to differentiation of the cross-correlation measurement; the resolved region $f_1 < 1/2t_{\mathrm{TOF}} < f_c$, in which the transfer function is an approximately linear scaling, and hence, the images of the causal and anti-causal Green's functions are resolved; and the unresolved region $f_c < 1/2t_{\mathrm{TOF}}$, in which the transfer function serves a band-limiting function, and hence, the images of the causal and anti-causal Green's functions cannot be resolved.

As a specific example of the intimate relationship between the spectral properties of the sensors and their separation, consider the Laplace domain transfer function of an *s*-wave seismogram[17]

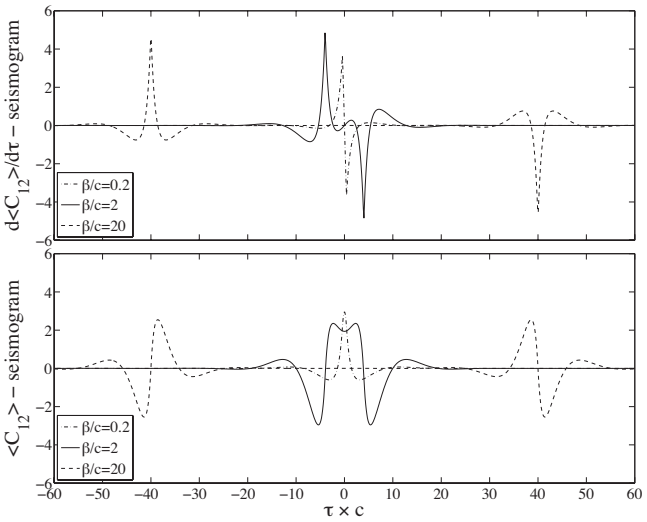


FIG. 2. The differentiated and undifferentiated cross-correlation measurements for the seismogram for $\beta/c=\{0.2,2,20\}$ and $D=10^{-4}$.

$$H_d = \frac{383.6s^3}{(s^2+0.5027s+0.1755)(s^2+1.257s+0.00487)}, \tag{8}$$

which has an impulse response function of

$$\begin{aligned} h_d(t) = &[-157.9\exp(-0.2514t)\cos(0.3351t) \\ &-107.8\exp(-0.2514t)\sin(0.3351t)]u(t) \\ &+541.5\exp(-0.6285t)u(t)[\cosh(0.6246t) \\ &-\sinh(0.6246t)]. \end{aligned} \tag{9}$$

The seismogram defined by Eq. (9) has $f_1 \approx 0.0564$ Hz and $f_c \approx 0.281$ Hz. Figure 2 shows plots of differentiated (upper) and undifferentated (lower) measurements that would be made with these seismograms for $\beta/c=\{0.2,2,2\}$. These correspond to the three regimes illustrated in Fig. 1. For $\beta/c=0.2$ the images of the causal and anti-causal Green's functions are smaller than for other values of β/c and not well located. For $\beta/c=2$, the images of the causal and anti-causal Green's functions are clearly resolved and accurately located. Finally, for $\beta/c=20$, the peaks are well located and look to be of a similar shape to those in the resolved region. To see why this might be deemed the differentiation regime, consider the undifferentiated measurements for $\beta/c=20$. These measurements exhibit doublets because each sensor individually acts as a differentiator, leading to a second-derivative.

It is also interesting to examine the relationship between sensor separation and sensor bandwidth results reported in other work, in the light of the rules of thumb presented in this paper. For example, in Ref. 10 it is reported that one experiment was carried out with sensor bandwidth 150–700 Hz, and a sensor separation of approximately 28 m. Noting that $1/2t_{\mathrm{TOF}}=c/4\beta=1500/56\approx 27$ Hz, this sensor is operating partly in the differentiation region.

In summary, this paper demonstrates the impact of the frequency response of sensors on estimation of the Green's function using ambient noise. The analysis leads to two rules-of-thumb concerning the relationship between cut-off frequencies and the distance between sensors. For a low-pass sensor, the Green's function can be resolved only if twice the time-of-flight between the sensors is greater than the time constant of the sensor. For a high-pass sensor, a differentiation effect occurs if the time-constant is much less than twice the time-of-flight between the sensors. The common case of a band-pass sensor can be thought of as the combination of the high and low-pass cases. These results have application in the choice of sensor bandwidths and separations in the design of measurement systems.

[1]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).
[2]N. M. Shapiro, M. Camillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface-wave tomography from ambient seismic noise," Science **307**, 1615–1618 (2005).
[3]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).
[4]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[5]P. Roux, S. A. Kuperman, and NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[6]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).
[7]J. Rickett and J. Claerbout, "Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring," The Leading Edge **18**, 957-960 (1999).
[8]J. N. Louie, "Faster, better: Shear-wave velocity to 100 meters depth from refraction microtremor arrays," Bull. Seismol. Soc. Am. **91**, 347–364 (2001).
[9]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).
[10]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).
[11]P. Roux and M. Fink, "Green's function estimation using secondary sources in a shallow water environment," J. Acoust. Soc. Am. **113**, 1406–1416 (2003).
[12]P. Roux, K. G. Sabra, W. A. Kuperman, and A. Roux, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2005).
[13]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 (L)," J. Acoust. Soc. Am. **113**, 2973–2976 (2003).
[14]N. M. Shapiro and M. Campillo, "Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise," Geophys. Res. Lett. **31**, L07614 (2004).
[15]R. Sneider, K. Wapenaar, and U. Wegler, "Unified Greens function retrieval by cross-correlation; connection with energy principles," Phys. Rev. E **75**, 036103 (2007).
[16]R. N. Bracewell, *The Fourier Transform and Its Applications*, 2nd ed. (McGraw-Hill, New York, 1986).
[17]Available online: http://www.geophys.uni-stuttgart.de/oldwww/seismometry/man_html/node10.html (Last viewed 4/12/08).

Comment on "Acoustic chaos in a duct with two separate sound sources" [J. Acoust. Soc. Am. 110, 120–126 (2001)] (L)

A. A. Castrejón-Pita
Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom

J. R. Castrejón-Pita
IfM, University of Cambridge, Mill Lane, Cambridge, CB2 1RX, United Kingdom

G. Huelsz
Centro de Investigación en Energía, Universidad Nacional Autónoma de México, A.P. 34, 62580 Temixco, Morelos, Mexico

A. Sarmiento-Galán
Instituto de Matemáticas, Universidad Nacional Autónoma de México, CP 62200 Chamilpa, Morelos, Mexico

(Received 24 March 2008; revised 21 July 2008; accepted 30 July 2008)

In a paper published in this journal in 2001 by Dong *et al.* [W. G. Dong, X. Y. Huang, and Q. L. Wo, J. Acoust. Soc. Am. **110**, 120–126 (2001)] it was claimed that acoustic chaos was obtained experimentally by the nonlinear interaction of two acoustic waves in a duct. In this comment a simple experimental setup and an analytical model is used to show that the dynamics of such systems corresponds to a quasiperiodic motion, and not to a chaotic one.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2978081]

PACS number(s): 43.25.Ts, 43.25.Zx, 43.58.Wc [ROC] Pages: 2702–2705

I. INTRODUCTION

The study of nonlinear systems has attracted the attention of scientists since the experiments of Jacques Hadamard were carried out in the early twentieth century. Since then, this phenomenon has been extensively studied by many authors, among others Poincairé, Kolmogorov, and Lorentz; a compilation of these and many other studies can be found elsewhere.[1] At present, the mathematical tools designed to analyze nonlinear dynamics and chaos are being applied to a great variety of dynamical systems; an example of this is the paper by Dong,[2] where an experimental setup consisting of two independently driven speakers placed at one end and on a side of a cylindrical duct and a microphone at the other end of the duct. The main objective of this paper was to define whether the resulting behavior of the interaction of two acoustic waves in a duct is chaotic or not. After analyzing an experimentally acquired time series and finding that (i) the largest Lyapunov exponent of the signal is positive, and (ii) the dimension of the reconstructed attractor is fractal, the authors concluded that the interaction of the two acoustic waves inside the duct is chaotic. In this comment, it is shown that the analysis presented in Dong's paper[2] is not sufficient to characterize the dynamics of such a system, since a quasiperiodic signal can be easily confused as a chaotic one if (a) the Lyapunov exponent analysis is not performed over an adequate time domain and (b) the error bars of the computed attractor fractal dimension include an integer value. Experimental data and a simple numerically generated signal are presented to support these affirmations.

Quasiperiodic behaviors are not difficult to produce or scarce in nature, but are difficult to characterize. In a simple case, a quasiperiodic motion can be thought of as a mixture of periodic motions with different incommensurable frequencies. As a result, a Fourier spectrum of a quasiperiodic signal consists of peaks at the values of the main fundamental frequencies and their linear combinations. The geometrical shape of its attractor in the phase space is a n-dimensional toroidal surface specified by the n fundamental frequencies. In fact, most of the nonlinear mathematical tools provide information to detect a quasiperiodic behavior. Detailed definitions and examples of Lyapunov exponents' analysis were given by Kantz,[3] but briefly, the Lyapunov exponent of a signal measures the rate of separation between nearby trajectories. If the distance between these trajectories does not change, it can be said that the dynamic response of the system is periodic and the largest Lyapunov exponent is zero. In contrast, if these trajectories diverge in time, exponentially to be precise, it could indicate that the system is chaotic and the largest Lyapunov exponent is positive. In a third case, if the system is quasiperiodic, these trajectories follow different paths but they would eventually approach each other only to diverge again and to repeat the process again and again. As a consequence, when a quasiperiodic behavior is suspected, the Lyapunov exponent analysis is required to be performed during times sufficiently large to identify these turn backs. In addition to this criterion, the time delay (denoted by J in Dong's paper[2]) used to calculate both the Lyapunov exponent and the dimension of the attractor D_c is also a critical variable to identify quasi-periodic behaviors. If it is taken too small, there would be almost no difference between the different elements of the delay vectors and this redundancy makes the vectors meaningless if the data is noisy and the

 © 2008 Acoustical Society of America

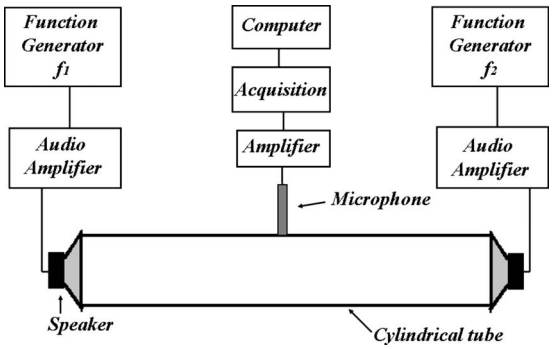


FIG. 1. Experimental arrangement.

variation of the signal during the lag J is less than the noise level.[4] On the other hand, if J is taken too large, then the different coordinates may be almost uncorrelated and the reconstructed attractor may become very complex, even if the "true" underlying attractor is a simple one. According to Swinney,[5] the method that estimates the value of J, using the first minimum of the time delayed mutual information is better and easier than the one that uses the autocorrelation.

In the next section, a simple experimental setup consisting of the same elements used in Dong's paper,[2] and which provides the time series analyzed below is presented. As expected, this data showed a Fourier spectrum with only the linear combination of the fundamental frequencies. Lyapunov exponents, return maps, and embedding dimensions were calculated for this data depicting a quasiperiodic motion and not a chaotic one. In addition, a simple quasiperiodic signal was created to test our dynamic-analysis battery; both, experimental and numerical results agree.

II. EXPERIMENTAL SETUP

As mentioned before, a simple experimental device, which basically consists of the same elements presented in Dong's paper,[2] was constructed and tested to prove our affirmations. Briefly, this setup consists of a cylindrical duct with speakers on both ends. The duct is a circular cylinder made of PVC, 60 cm long, 8 cm in diameter, and a 3 mm-wide wall. Two speakers PROAM model SPK-350 of 8 Ω and 75 W were placed at each one of the two ends of the duct. For each speaker, a function generator SRS was connected to an audio amplifier (ONKYO, with total harmonic distortions <0.001) to produce the input-speaker voltage signals. A Bruel & Kjaer microphone with its Measuring Amplifier (2610 type) was installed on the side of the duct to detect the pressure signal. The output of the microphone amplifier was measured and recorded by an acquisition system HP 3206A connected to a computer using HP-VEE software.

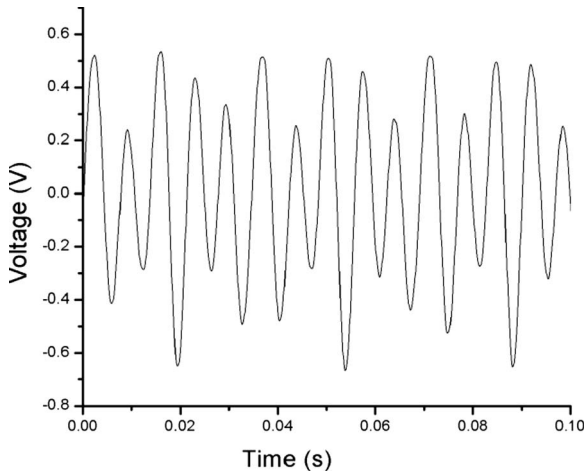


FIG. 2. Temporal signal with $f_1=88$, $f_2=145$ Hz, and $V_{in}=2$ V.

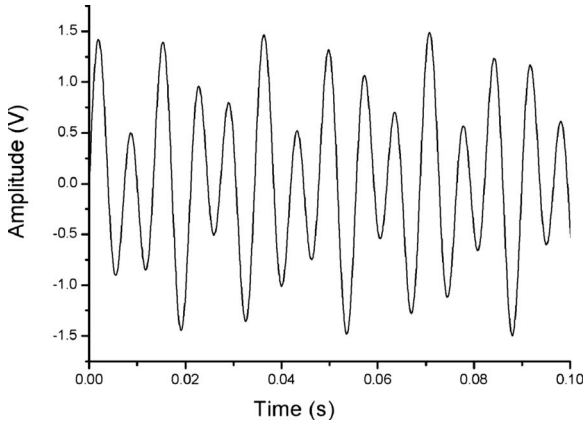


FIG. 3. Numerical signal with $f_1=88$ and $f_2=145$ Hz.

Each speaker was driven by a pure sinusoidal signal of one of the two incommensurable frequencies: $f_1=88$ or $f_2=145$ Hz, both of them with an amplitude of $V_{in}=2$ V (Fig. 1). A typical signal from the microphone is shown in Fig. 2; Fig. 3 shows a numerical signal used to mimic this dynamics which is simply a linear combination of two pure sinusoidals with frequencies f_1 and f_2.

III. ANALYSIS AND RESULTS

There are many criteria to classify dynamical systems, the most common are: the return map (commonly known as the experimental Poincaré section), Fourier spectrum, the phase space embedding dimension, the sign of the largest Lyapunov exponent, and the time delayed reconstructed attractor and its dimension.[3,6–8] Our analyzes were performed in terms of these time series analysis using the algorithms provided by Kantz.[8]

The step was to find out if the temporal signal was periodic or not via the return map (constructed by plotting the local extremal values of the time series as a function of the immediately preceding extrema). The return map can also help in the distinction between periodic and chaotic dynamics; it is, however, a qualitative test only and more careful analysis should be performed before trying to draw any conclusion. Qualitatively, there are several possibilities. The plot may consist of single points (whose number indicates the periodicity of the system) or it may be a simple (thin) closed curve that indicates a probable quasiperiodic dynamics. On the other hand, if the plot is an open (usually thick) curve, then the chances of having a chaotic dynamical system increase.

The return maps obtained for the signals are shown in Figs. 4 and 5, and correspond to a closed (period-one) curve. As mentioned above, this is a distinctive signature of a quasiperiodic motion.

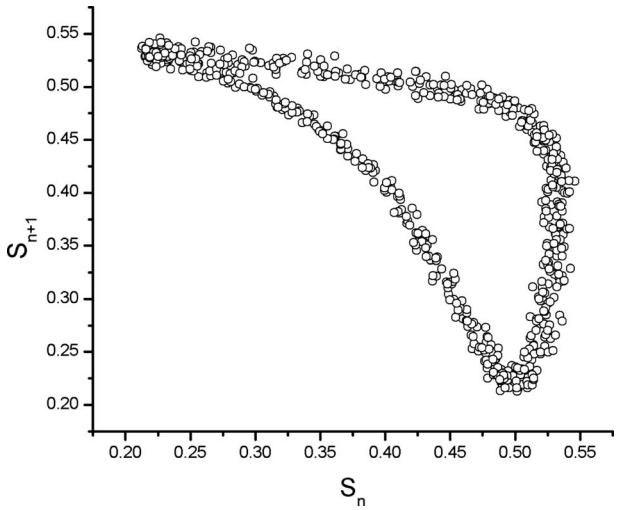


FIG. 4. Return map of the temporal signal in Fig. 2.

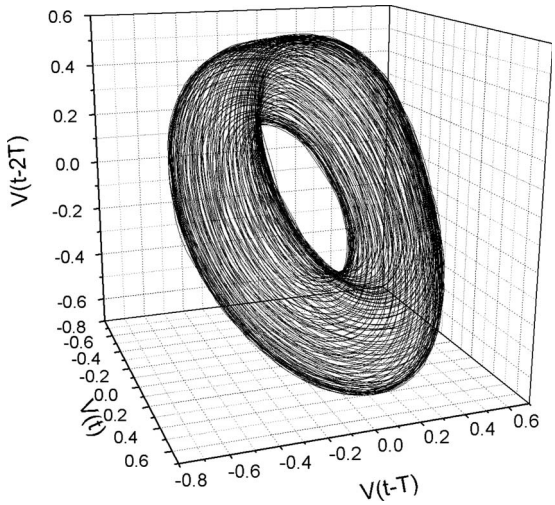


FIG. 6. Reconstructed attractor of the temporal signal in Fig. 2, and $J=2.1$ ms, $D_e=3$, and $D_c=2.02$.

To find the attractor we reconstructed the state space of the system using the well known method of time-delayed graphs. There are two important factors for this process: the time delay, J, used for the reconstruction and the embedding dimension, D, of the phase space where the attractor will be reconstructed. An incorrect choice of the embedding dimension for the reconstruction of the attractor could imply a wrong interpretation of the behavior of the dynamical system. An underestimate of the true dimension could show self-intersections or false periodicity due to the projection of a higher dimensional object. For the calculation of the time delay, an algorithm that uses the statistical ideas of the average mutual information was employed; as presented by Kantz.[8] An optimal value of $J=2.1$ ms is obtained using this method. To compute a good estimate of the embedding dimension, D_e, we used the false neighbors method which measures the Euclidean distance between two neighboring points of the attractor assuming it is embedded in a D-dimensional space and compares this value to the one obtained assuming a $(D+1)$-dimensional embedding space. If going from dimension D to $D+1$ unprojects away the two points by showing a clear increment of the Euclidean distance, this then shows that they were false neighbors. The dimension that has the lowest percentage of false neighboring points gives a good estimate for the true embedding dimension. Again, we used the algorithm provided by Kantz.[8]

The reconstructed attractors, using the values obtained via the time series analysis, are presented in Figs. 6 and 7, and correspond to tori. Using the Grassberger–Procaccia algorithm, the dimension of the attractor (D_c) was calculated to be $D_c=2.02\pm0.09$, a value which lies on a range where no conclusions can be drawn as it could indicate an integer dimension. In Dong's paper,[2] D_c varies from 1.98 to 2.39. As this range surrounds an integer value, it should not be used to conclude a chaotic dynamics. In fact, there is another point of concern in Dong's paper[2] as their text-reported values of D_c do not correspond with the linear fits shown in their graphs; Figs. 12 and 13 of Ref. 2 show values of $D_c^{J=4}=2.35$ and $D_c^{J=10}=2.33$, respectively. However, later in the text it is stated that $D_c^{J=4}=2.15$ and $D_c^{J=10}=2.5$.

The Kantz's algorithm[3,8] was used to calculate the largest Lyapunov exponent. The value of this exponent is obtained by computing the slope of a well-defined monotoni-

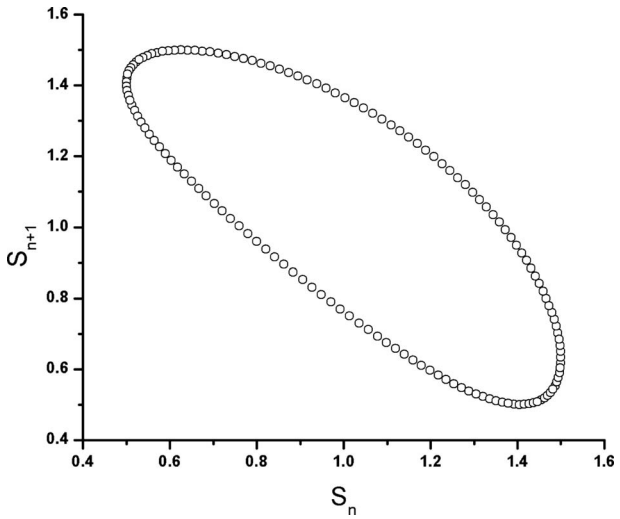


FIG. 5. Return map of the numerical signal in Fig. 3.

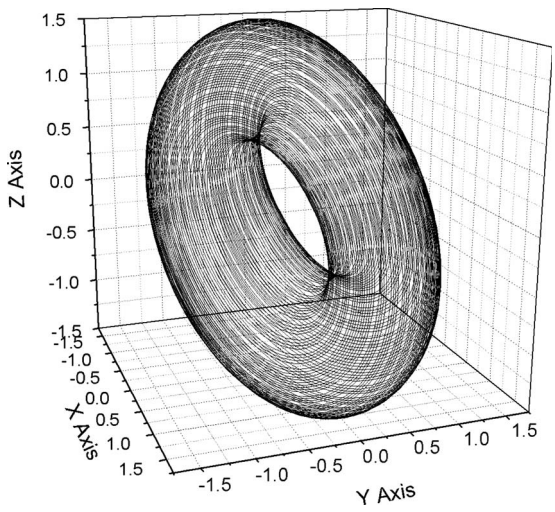


FIG. 7. Reconstructed attractor of the numerical signal in Fig. 3, and $J=2.1$ ms, $D_e=3$, and $D_c=1.95$.

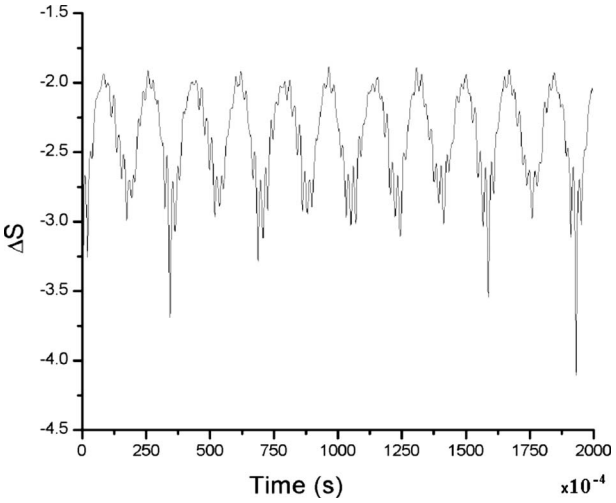


FIG. 8. ΔS for the experimental time series with $f_1=88$ Hz, $f_2=145$ Hz, $V_{\text{in}}=2$ V, $J=2.1$ ms, and $D_e=3$.

cally growing plot of the divergence of nearby trajectories in the phase space (in logarithmic scale because the algorithms assumed an exponential growth of this divergence). Our results are presented in Figs. 8 and 9. It is important to notice that the Lyapunov exponent (the slope of ΔS versus time), when calculated for reasonably large periods, is undefined. In particular, the divergence of adjacent trajectories in time, Figs. 8 and 9, is oscillatory for both the experimental and the numerical signals presented in this work (this periodic behavior is evident along the entire length of the data). This behavior is also observable in Figs. 10 and 11 in Dong's paper,[2] although there ΔS [ln(divergence)] is shown for a short period of time and only one oscillation is present. In addition, from the reconstructed attractor, the return map, the fractal dimension, and the Fourier spectrum of our experimental data we can conclude that there is not evidence of a chaotic behavior. First, the reconstructed attractor clearly represents a two-dimensional tori. Second, the return map is a closed and well defined curve. Third, the Fourier spectrum (not shown) peaks only at the original frequencies, their harmonics, and linear combinations of them. The lack of a continuous component in a frequency power spectrum is a hallmark of a periodic or quasiperiodic dynamics.

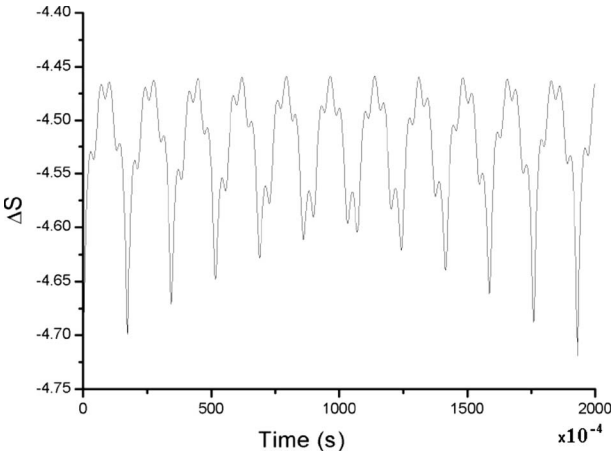


FIG. 9. ΔS for the numerical signal with $f_1=88$ Hz, $f_2=145$ Hz, $J=2.1$ ms, and $D_e=3$.

There is an artifact on the methods used to calculate the divergence of trajectories when they are applied to periodic (with more than one fundamental frequency) and quasiperiodic signals. Those methods are designed to look for an exponential divergence of nearby trajectories, and hence, the results are rather spurious when the divergence is not exponential. Interestingly, if one estimates the period of oscillation in graphs 10 and 11 in Dong's paper,[2] or Figs. 8 and 9 in this work, it turns out to be nothing else than the difference between the fundamental frequencies of the systems, i.e., 730 Hz in the original paper, and 57 Hz in our case (145–88 Hz). This tells us then, that if one thinks of two nearby trajectories in phase space, they never touch each other but get very close in regions near the "hole" of the torus and are relatively apart in the "opposite" regions of the same torus. This qualitatively explains the oscillatory shape of the plots presented in Figs. 8 and 9 in this work or in Figs. 10 and 11 of the original paper.

IV. CONCLUSIONS

From our analysis, we can conclude that the dynamics reported in Dong's paper[2] is quasiperiodic and not chaotic. This conclusion is based on the fact that the results in the calculations of the Lyapunov exponent were misinterpreted: incorrectly calculated to show an apparent positive value. This was presented as the strongest evidence for chaos. If the calculation would had been carried out for longer time span, it would had shown the oscillatory behavior in the divergence of the trajectories [ln(divergence) in Dong's paper,[2] or ΔS in this paper] from where the Lyapunov exponent is calculated, in the same way that it was presented here. To avoid this confusion, we propose the calculation of the largest Lyapunov exponent to be performed for longer times as this would allow the observation of a clear oscillatory behavior. Our experiments were also carried out all over the speaker optimum operational range, from 2 to 25 V, with similar results; no signatures of chaotic behavior were found.

ACKNOWLEDGMENTS

This work has been supported by CONACyT project 25116. A.A.C-P. thanks the UK Natural Environment Research Council for support.

[1]K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos: An Introduction to Dynamical Systems*, 1st Ed. (Springer, New York, 2000).
[2]W. G. Dong, X. Y. Huang, and Q. L. Wo, "Acoustic chaos in a duct with two separate sound sources," J. Acoust. Soc. Am. **110**, 120–126 (2001).
[3]H. Kantz and T. Shreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 2002).
[4]H. Kantz, "A robust method to estimate the maximal Lyapunov exponent of a time series," Phys. Lett. A **185**, 77 (1994).
[5]A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Phys. Rev. A **33**, 1134 (1986).
[6]H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. Sh. Tsimring, "The analysis of observed chaotic data in physical systems," Rev. Mod. Phys. **65**, 1331–1392 (1993).
[7]R. C. Hilborn, *Chaos and Nonlinear Dynamics* (Oxford University Press, Oxford, 1994).
[8]R. Hegger, H. Kantz, and T. Shreiber, "Practical implementation of nonlinear series methods: The TISEAN package," Chaos **9**, 413 (1999).

Comment on "Cepstrum pitch determination" [J. Acoust. Soc. Am. 41, 293–309 (1967)] (L)

Arturo Camacho[a)]
Computational NeuroEngineering Laboratory, University of Florida, Gainesville, Florida 32611

(Received 16 May 2008; revised 25 August 2008; accepted 26 August 2008)

In a paper by A. Michael Noll [J. Acoust. Soc. Am. **41**, 293–309 (1967)], the use of the cepstrum was proposed to determine the pitch of a signal. This paper shows that such a method does not work on periodic signals. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2988293]

PACS number(s): 43.66.Hg [EJS] Pages: 2706–2707

Noll (1967) proposed the use of the cepstrum to determine the pitch of a signal. The cepstrum $C(\tau)$ of a signal $f(t)$ is defined as the power spectrum of the logarithm of its power spectrum

$$C(\tau) = \left\{ \int_0^\infty \log|F(\omega)|^2 \cos(\omega t) d\omega \right\}^2, \tag{1}$$

where $F(\omega)$ is the Fourier transform (FT) of $f(t)$, with ω measured in radians.

This approach has a serious limitation that makes it fail in what we expect to be the best-case scenario: a periodic signal. If $f(t)$ is periodic with period $T_0 = 2\pi/\omega_0$, it has to be of the form

$$f(t) = \sum_{k=0}^{\infty} c_k \cos(k\omega_0 + \phi_k), \tag{2}$$

and the logarithm of its power spectrum must be

$$\log|F(\omega)|^2 = \begin{cases} \log c_k^2, & \omega = k\omega_0, \\ -\infty, & \text{otherwise}. \end{cases} \tag{3}$$

Since this function is nonfinite over some open sets of its domain, the integral in Eq. (1) does not exist. To avoid this, Noll suggested replacing the minus infinity in Eq. (3) with a (presumably small) finite value, (i.e., a negative number with a large absolute value). Alternatively, other researchers have suggested adding a small positive constant ε to the argument of the logarithm

$$C'(\tau) = \left\{ \int_0^\infty \log[|F(\omega)|^2 + \varepsilon] \cos(\omega t) d\omega \right\}^2. \tag{4}$$

Although both of these modifications solve the problem under consideration, there is yet another problem they do not solve: Cepstrum gives too much "weight" to low-energy spectral components. An example will illustrate why this is the case. Consider a periodic pulse train with period 10 ms, as shown in Fig. 1(a). Its spectrum [Fig. 1(b)[1]] and its cepstrum (not shown in the figure) are also periodic pulse trains, the latter with same period as the signal (10 ms). If notches of magnitude 500 are introduced in the log-spectrum of the signal at multiples of 70.7 Hz[2] [Fig. 1(c)], no auditory difference should be perceived, because 500 in natural base corresponds to approximately 2000 dB, and the range of human auditory perception is only 120 dB. However, the cepstrum of this spectrally notched signal [Fig. 1(d)] is very different from the cepstrum of the original signal: All its major peaks are located at multiples of 14.1 ms rather than at multiples of 10 ms, suggesting a pitch of 70.7 Hz. A solution to this problem is to make $\varepsilon = 1$ in Eq. (4), which causes the weight of the spectral components approach zero as their energy approaches zero. However, this trick makes the cepstrum sensible to the scaling of the signal (if the scale is large, the weighted spectrum approaches a logarithmic behavior; if the scale is small, the weighted spectrum approaches linearity).

We conclude this letter with a comment related to the deconvolving capability of the cepstrum, which is probably the reason why cepstrum is so widely used in speech processing. A speech signal $f(t)$ is formed by the convolution of an excitation signal $e(t)$ and the combined impulse response $h(t)$ of the vocal tract and the radiation at the lips. Their respective FTs, $F(\omega)$, $E(\omega)$, and $H(\omega)$, hold the following relation:

$$|F(\omega)| = |E(\omega)||H(\omega)|. \tag{5}$$

If $e(t)$ is (quasi) periodic with period T_0, $|E(\omega)|$ has (major) peaks at multiples of $2\pi/T_0$, while $|H(\omega)|$ is usually smooth, and serves as an envelope for $|E(\omega)|$. Squaring them and taking the logarithm of their product does not change these properties, but conveniently converts this product into a sum

$$\log|F(\omega)|^2 = \log|E(\omega)|^2 + \log|H(\omega)|^2. \tag{6}$$

If this sum is finite for all ω, its FT exists, and because of the linearity of the FT, it can be written as

$$\int_0^\infty \log|F(\omega)|^2 \cos(\omega t) d\omega = \int_0^\infty \log|E(\omega)|^2 \cos(\omega t) d\omega + \int_0^\infty \log|H(\omega)|^2 \cos(\omega t) d\omega. \tag{7}$$

The FT of $\log|E(\omega)|^2$ has (major) peaks at multiples of T_0, and because of its slow varying nature, the FT of $\log|H(\omega)|^2$ has most of its energy at low *quefrencies* (the unit used in the cepstrum domain), usually much lower than T_0. Henceforth,

a)Electronic mail: acamacho@cise.ufl.edu

 © 2008 Acoustical Society of America

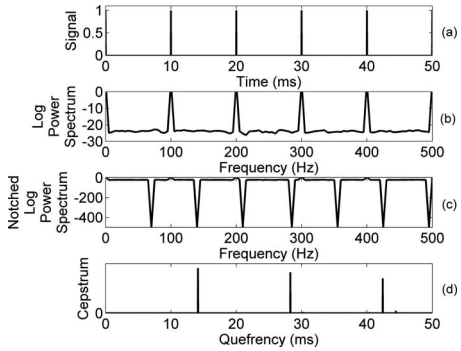


FIG. 1. (a) First 50 ms of a 200 ms pulse train with period 10 ms and sampling frequency 51.2 kHz. (b) First 500 Hz of the spectrum of the signal, computed using a rectangular window with same length as the signal (200 ms). (c) Same spectrum as in (b) but with notches at multiples of 70.7 Hz. (d) First 50 ms of the notched-spectrum-based cepstrum (the value at zero was set to zero because it is irrelevant for pitch determination).

it is usually the case that these FTs are separable through *liftering* (filtering in the cepstrum domain). However, setting a lower bound on the logarithm (as suggested by Noll) or adding a constant to the argument of the logarithm [as in Eq. (4)] excludes the possibility of converting the logarithm of the product into a sum of logarithms, as done in Eq. (6). Consequently, the deconvolving property of cepstrum can be used only if the power spectrum of the signal has no zeros, which allows dispensing with the use of the modifications to the cepstrum discussed here.

[1]The log-spectrum in the figure is finite everywhere not because of the use of the modifications under discussion but because of the use of finite-precision arithmetic.

[2]This value corresponds to the irrational number $100/\sqrt{2}$. It was chosen to be irrational to avoid any matches between its harmonics and the harmonics of 100 Hz.

Individualized head-related transfer functions based on population grouping (L)

Song Xu, Zhizhong Li,[a] and Gavriel Salvendy
Institute of Human Factors and Ergonomics, Department of Industrial Engineering, Tsinghua University, Beijing, People's Republic of China

(Received 8 May 2008; revised 7 August 2008; accepted 15 August 2008)

A method is proposed to divide a population into different groups for partial individualization of head-related transfer functions (HRTFs). Borrowing the basic idea in sizing system design, factor analysis is used to identify the most representative measurements which are then in a case study used to group the population. The comparison between the group mean HRTFs and the population mean HRTFs shows that the group mean HRTFs could greatly reduce spectral distortion at most sampled positions. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2982398]

PACS number(s): 43.66.Pn, 43.66.Qp [JCM] Pages: 2708–2710

I. INTRODUCTION

Previous studies have addressed the need of head-related transfer function (HRTF) individualization[1–7] since nonindividualized HRTFs may cause errors and lower a listener's perception performance of a spatial virtual sound. HRTFs can be indirectly individualized by using a certain level of correlation between physical features and HRTFs.[8] A HRTF could be modeled as a linear combination of several basic and representative spectral shapes.[9] Based on it, a method was developed to individualize HRTFs using listeners' physical features.[8] Several studies have emphasized the importance of anthropometric measurements of external ears, head, and torso in sound localization.[10–12] Along with a case study, a method based on population grouping by anthropometric measurements is proposed to partially individualize HRTFs in this paper.

II. METHOD AND RESULTS

HRTFs could be possibly individualized by several limited anthropometric measurements of a listener.[8] It motivates us to divide a population into several groups in each of which the listeners have similar HRTFs by borrowing the basic method of anthropometric sizing system design. Thus, the sonic products based on this idea could be chosen by a listener according to his/her "size" similar to choosing a pair of shoes for himself/herself.

Four steps can be carried out to obtain the individualized HRTFs by population grouping. First, some anthropometric measurements related with HRTFs are screened; then, the most representative measurements are detected from the related measurements; third, the population is divided into several groups according to the representative measurements; finally, the average HRTFs within each group are calculated as the individualized HRTFs. The details of the method are described as following along with a case study.

A. Screening anthropometric measurements

Some anthropometric measurements of the head, ears, and torso play a significant role in a listener's perception of a spatial virtual sound. The anthropometric data of 1774 male subjects in the U.S. Army Anthropometric Survey (ANSUR) database[13] were analyzed as a case. Seventeen measurements, which were hypothesized to be HRTFs related, were chosen from the database: stature, bideltoid breadth, ear length, sitting height, bizygomatic breadth, chest circumference below breast, cervicale height, chest depth, ear breadth, ear length above tragion, ear protrusion, head breadth, head circumference, head length, neck height lateral, shoulder circumference, and suprasternale height.

B. Factor analysis

Among the measurements related with HRTFs, some of them are strongly correlated with each other. Factor analysis is a statistical method to explain variability among observed random variables in terms of fewer unobserved random variables (factors). Here it was used to identify the most representative measurements from the 17 measurements. Three representative factors were identified. Factor 1, the vertical factor which represents vertical measurements, mainly includes stature, cervicale height, suprasternale height, and neck height lateral. Factor 2, the horizontal factor which represents horizontal measurements, mainly includes chest circumference below breast, bideltoid breadth, chest depth, and shoulder circumference. Factor 3, the ear factor which represents the size of ears, includes ear length and ear length above tragion. Bideltoid, ear length, and stature were chosen as the representative measurements of the three factors since they had the largest three eigenvalues and totally explained 62% variance of the data (37%, 15%, and 10%, respectively).

C. Population grouping

The three representative measurements can then be used to group the population. In the case study, the range of each representative measurement was divided into three equal intervals (numbered as 1, 2, and 3 in Table I). Theoretically, the population can be divided into 27 groups. The grouping results of the 1174 male subjects in the ANSUR database are listed in Table I. There are eight groups (marked with letter[b]) to each of which more than 3% male subjects in the ANSUR

[a]Electronic mail: zzli@tsinghua.edu.cn

 © 2008 Acoustical Society of America

TABLE I. The results of grouping population.

Group level		No. of subjects (ANSUR)	No. of subjects (CIPIC)
No.	(*S*-*B*-*E*)[a]		
1	1-1-1	37	—
2	1-1-2	61[b]	—
3	1-1-3	3	—
4	1-2-1	27	—
5	1-2-2	87[b]	1
6	1-2-3	3	—
7	1-3-1	0	—
8	1-3-2	1	—
9	1-3-3	1	—
10	2-1-1	48	2
11	2-1-2	196[b]	6
12	2-1-3	14	—
13	2-2-1	132[b]	3
14	2-2-2	865[b]	7
15	2-2-3	104[b]	—
16	2-3-1	5	—
17	2-3-2	72[b]	—
18	2-3-3	13	—
19	3-1-1	2	1
20	3-1-2	5	—
21	3-1-3	1	—
22	3-2-1	4	—
23	3-2-2	59[b]	—
24	3-2-3	12	—
25	3-3-1	1	—
26	3-3-2	15	1
27	3-3-3	6	—
Sum		1774	21

[a]*S*=stature (height), *B*=bideltoid breadth (shoulder width), *E*=ear length (pinna height).
[b]Groups with more than 3% subjects.

database belong. The CIPIC database, which was developed by the Center for Image Processing and Integrated Computing (CIPIC) Interface Laboratory of the University of California at Davis, contains HRTFs and anthropometric measurements for 21 male subjects. In the database, the three representative measurements have different names, i.e., shoulder width (for bideltoid), pinna height (for ear length), and height (for stature). By comparing the values of the three representative measurements with the group intervals, which group each subject in the CIPIC database belongs to can be determined. The results are also given in Table I.

D. HRTF individualization and validation

The partially individualized HRTFs of a group, which is called "group mean HRTFs," can be obtained by averaging HRTFs within the group. A "population mean HRTF" refers to the average value of all subjects' HRTF (across groups) at a specified position. The group mean HRTFs of groups 11, 13, and 14 in Table I were calculated according to the HRTFs data of the associated 16 subjects at 1250 positions in the CIPIC HRTF database.[14] They were compared with the population mean HRTFs of the CIPIC database using spectral distortion (SD). A SD score is an objective criterion that has been widely used to evaluate speech quality before it was introduced to evaluate HRTF estimation. It has been used to describe the errors of HRTF estimation using physical features.[8] A SD score at position (φ, θ) is given by

$$\mathrm{SD}_{\varphi,\theta} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}[h_{\varphi,\theta}(f_j) - \hat{h}_{\varphi,\theta}(f_j)]^2}\ (\mathrm{dB}),$$

where $h_{\varphi,\theta}(f_j)$ is the measured HRTF of the jth frequency at position (φ, θ) and $\hat{h}_{\varphi,\theta}(f_j)$ is the population or group mean HRTF (N is the number of frequency, φ=azimuth, θ=elevation, HRTF in the form of log-magnitude with dB unit).[8]

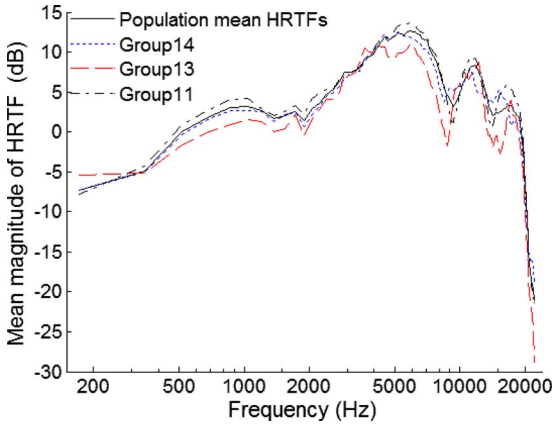


FIG. 1. (Color online) Mean magnitudes of population mean HRTFs and group mean HRTFs for the left ears of the male subjects in the CIPIC database at position (−80°, 0°).

Figure 1 shows the population mean HRTFs and the group mean HRTFs of groups 11, 13, and 14 at position (−80°, 0°) of the left ear for the male subjects. It can be seen that the group mean HRTFs are different from the population mean HRTFs. The descriptives of the SD scores of the group and population mean HRTFs at all the 1250 positions are given in Table II. For group mean HRTFs, the means of SD scores of groups 11, 13, and 14 are 5.034, 4.742, and 4.980 dB, respectively. It can be found that compared with the population mean HRTFs, the group mean HRTFs have

TABLE II. The descriptives of the SD scores when using the group and population mean. PH=population mean HRTF; GH=group mean HRTF.

	Group 11		Group 13		Group 14	
Statistic	$\mathrm{SD}_{\mathrm{PH}}$	$\mathrm{SD}_{\mathrm{GH}}$	$\mathrm{SD}_{\mathrm{PH}}$	$\mathrm{SD}_{\mathrm{GH}}$	$\mathrm{SD}_{\mathrm{PH}}$	$\mathrm{SD}_{\mathrm{GH}}$
Mean	5.370	5.034	6.239	4.742	5.233	4.980
Standard	0.881	0.876	1.346	0.662	0.814	0.768
Maximum	9.034	9.164	13.658	7.628	9.323	9.083
Minimum	3.354	2.786	3.713	2.705	3.118	2.965

lower means of SD scores in the three groups by 6.3%, 22.5%, and 4.7%, respectively. By using the group mean HRTFs, for group 11, SD is reduced by at least 5% at 741 positions (about 60% of 1250 positions); for group 13, SD is reduced by at least 10% at 1174 positions (about 94% of 1250 positions); for group 14, SD is reduced by at least 5% at 552 positions (about 44% of 1250 positions).

III. CONCLUSIONS

A method to partially individualize HRTFs based on population grouping is proposed. By this method, three factors (horizontal, vertical, and ear) related with HRTFs were identified to divide the population. A case study is presented to demonstrate the grouping method and SDs when using the group mean HRTFs. It was found that compared with the population mean HRTFs, the group mean HRTFs could greatly reduce the SDs at most sampled positions. The new idea makes it possible to partially individualize HRTFs using a few anthropometric measurements and might be used to design sonic products with "sizes."

[1]E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123 (1993).

[2]H. Møller, M. F. Sørensen, D. H. Hammershgi, and C. B. Jensen, "Head related transfer function of human subjects," J. Audio Eng. Soc. **43**, 300–321 (1995).

[3]H. Møller, M. F. Sørensen, and C. B. Jensen, "Binaural technique: Do we need individual recordings," J. Audio Eng. Soc. **44**, 451–469 (1996).

[4]P. J. Minnaar, S. K. Olesen, F. Christensen, and H. Møller, "Localization with binaural recordings from artificial and human heads," J. Audio Eng. Soc. **49**, 323–336 (2001).

[5]J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492 (1999).

[6]J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," J. Acoust. Soc. Am. **106**, 1493–1510 (1999).

[7]J. C. Middlebrooks, E. A. Macpherson, and Z. A. Onsan, "Psychophysical customization of directional transfer functions for virtual sound localization," J. Acoust. Soc. Am. **108**, 3088–3091 (2000).

[8]T. Nishino, N. Inoue, K. Takeda, and F. Itakura, "Estimation of HRTFs on the horizontal plane using physical features," Appl. Acoust. **68**, 897–908 (2007).

[9]D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91**, 1637–1647 (1992).

[10]J. Blauert, *Spatial Hearing* (MIT, Cambridge, MA, 1997), pp. 78–89.

[11]H. L. Han, "Measuring a dummy head in search of pinna cues," J. Audio Eng. Soc. **42**, 15–37 (1994).

[12]V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," J. Acoust. Soc. Am. **109**, 1110–1122 (2001).

[13]C. Clauser, I. Tebbetts, B. Bradtmiller, J. McConville, and C. C. Gordon, "Measurer's handbook: U.S. army anthropometric survey," Anthropology Research Project, Inc., 1988.

[14]V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, 2001, pp. 99–102.

A re-evaluation of the relation between physiological channel interaction and electrode pitch ranking in cochlear implants (L)

Michelle L. Hughes
Boys Town National Research Hospital, Lied Learning and Technology Center, 425 North 30th Street, Omaha, Nebraska 68131

(Received 29 May 2008; revised 28 August 2008; accepted 29 August 2008)

Spatial excitation patterns in cochlear implant users can be measured with the electrically evoked compound action potential (ECAP). This study examined whether the relative separation of ECAP excitation patterns for two electrodes was correlated with the ability to discriminate those electrodes on the basis of pitch. Significant correlations were found for nine of the ten subjects. Electrodes with significant relative overlap of ECAP spatial excitation patterns were generally more difficult to distinguish on the basis of pitch. Pitch-ranking ability and overlap of ECAP patterns were both affected by the relative separation between electrodes in each pair. With increased separation between electrodes, pitch ranking improved significantly, and ECAP spatial excitation patterns showed significantly less overlap.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2990710]

PACS number(s): 43.64.Me, 43.64.Pg, 43.66.Fe [BLM] Pages: 2711–2714

I. INTRODUCTION

Present cochlear implant (CI) arrays consist of 12–22 electrodes spaced along the length of the cochlea. Different pitch percepts are achieved through stimulation at different locations along the array. For CI users, the ability to discriminate electrodes on the basis of pitch is likely influenced by the spatial overlap of stimulated neural populations. Spread of excitation (SOE) patterns for CI electrodes can be measured with the electrically evoked compound action potential (ECAP) using a forward-masking paradigm (Abbas *et al.*, 2004; Cohen *et al.*, 2003; Eisen and Franck, 2005; Hughes and Abbas, 2006). Hughes and Abbas (2006) investigated whether electrodes with more restricted SOE patterns were more easily discriminated on the basis of pitch. Widths of SOE patterns were compared with slopes of electrode pitch-ranking functions. It was hypothesized that steeper slopes (better pitch ranking) would be correlated with narrower SOE patterns. No significant correlations were found for individual or group data. It was noted that the SOE width measure might not adequately capture other aspects of the function that may be predictive of electrode pitch ranking, such as relative location of the pattern edges or the overall spread of the function (e.g., McKay *et al.*, 1999).

In this letter, data from the Hughes and Abbas (2006) study were re-analyzed to examine the relation between relative separation of SOE functions and electrode pitch-ranking accuracy for pairs of electrodes. There were three notable differences with the new analysis: (1) Comparisons of SOE and pitch-ranking measures were made for electrode pairs, whereas the initial study used single electrodes; (2) The entire SOE function was used to measure overlap, whereas the initial study used only the width at 75% of the normalized amplitude; (3) Pitch-ranking percent correct was used instead of the slope of the psychometric function. It was hypothesized that electrodes with greater overlap of SOE patterns would be more difficult to distinguish on the basis of pitch.

II. METHODS

Data from the ten adult CI users from Hughes and Abbas (2006) were re-evaluated ($N=5$ Nucleus 24M, $N=5$ Nucleus 24R[CS]). Mean age at implant was 59 years (range: 35–72), mean duration of CI use was 2.6 years (range: 5 months–5 years), and mean duration of deafness prior to implantation was 9.1 years (range: 2 months–48 years). For additional subject demographics, see Table I of Hughes and Abbas (2006).

Behavioral threshold and maximum comfort levels were determined using clinical programming procedures and software (WinDPS, R116, build 445) via a SPrint speech processor and processor control interface (PCI). The stimulus was a 250-pps, 500-ms pulse train of 25-μs/phase biphasic current pulses presented in monopolar mode (re: MP1). Stimuli were loudness balanced across electrodes at 50% and 100% of the dynamic range. Subjects reported all electrodes to be in tonotopic order based on subjective pitch of stimuli swept across the array. Stimulation levels were determined from these behavioral dynamic ranges, as described below.

ECAPs were obtained using the Neural Response Telemetry software (v. 3.0) via a SPrint speech processor and commercial interface (PCI or portable programming system). Default stimulus and recording parameters were used: 80 Hz rate, 25 μs/phase pulse width, 60 dB gain, MP1 stimulus reference electrode, and MP2 recording reference electrode. A forward-masking paradigm was used to measure spatial excitation patterns and to separate the ECAP response from stimulus artifact [see Hughes and Abbas (2006) for further details]. Briefly, the probe pulse was fixed on one electrode and the location of the masker was varied across the array.

The resulting ECAP amplitude represents the overlap between masker and probe, where the largest amplitude should occur when masker and probe are delivered to the same electrode, with progressively smaller amplitudes for greater separations between masker and probe. Current levels were fixed at 80% of the behavioral dynamic range for each electrode, except for M24 and M50, who required levels at 90% for measurable ECAPs. ECAP recordings were an average of 50–200 sweeps, recorded 2–3 electrode positions apical to the probe electrode. ECAP amplitudes were measured as the difference between the leading negative peak (N1) and the following positive peak or plateau (P2).

For the re-analysis, ECAP SOE functions were normalized to the single highest amplitude of all ECAPs within each subject to allow comparisons within and across subjects. In Hughes and Abbas (2006), SOE functions were normalized separately for each probe electrode. The present approach preserved the relative amplitude differences across functions within each subject to allow comparisons of different pairs of electrodes.

For each electrode pair that was tested for pitch ranking, the respective SOE functions were compared and quantified. The amount of separation between the two SOE functions, henceforth termed ECAP separation index, was calculated as the absolute value of the difference in normalized amplitude, summed across all masker electrodes

$$\sum_{i=1}^{22} |a_{xi} - a_{yi}|,$$

where a_x and a_y represent the normalized amplitudes of the two ECAP functions for probe electrodes x and y at each masker electrode i. Two examples illustrating this method are shown in Fig. 1 for M35b. In Fig. 1(a), ECAP SOE functions for probe electrode 9 (P9) and P11 are shown; Fig. 1(b) shows SOE functions for P9 and P13. The peak of each function typically occurs at the probe electrode. Therefore, relative to the point where the functions cross, amplitudes are higher for the more basal probe (P9) on the basal side and for the more apical probe (P11 or P13) on the apical side. Vertical lines connecting the symbols represent the amplitude difference at each masker electrode. The ECAP separation index was 1.7 for P9 versus P11 and 4.2 for P9 versus P13. Thus, a larger ECAP separation index indicates greater separation between SOE functions.

The stimulus for pitch ranking was the same as that used to determine behavioral dynamic range. Pitch ranking was evaluated using a two-interval (separated by 500 ms), two-alternative, forced-choice procedure in which the subject indicated whether the second sound was higher or lower in pitch relative to the first. Stimulus levels were jittered between 70%, 80%, and 90% of the behavioral dynamic range to reduce the effects of loudness cues[1]. For the present analysis, the existing data set was examined to identify pairs of electrodes in which each electrode occurred in the first and second interval an equal number of times[2]. This resulted in as few as three electrode pairs (subject R13) to as many as 39 pairs (M24) across subjects. A total of 36 comparisons were obtained for each electrode pair: nine comparisons per block

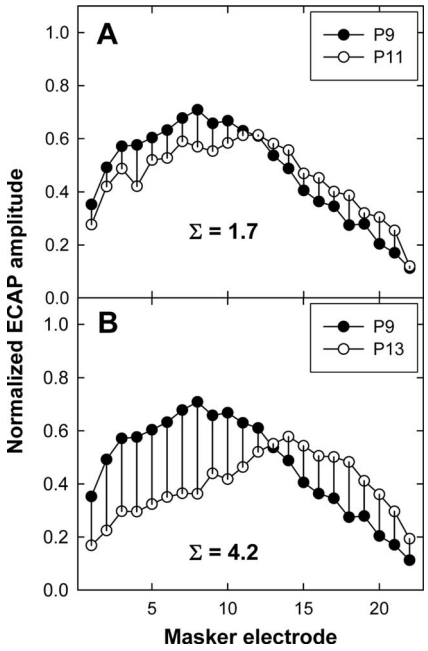


FIG. 1. Comparison of normalized ECAP SOE patterns for probe electrode P9 vs P11 (panel A) and P9 vs P13 (panel B) for subject M35b. Vertical bars indicate the difference in normalized amplitude at each masker electrode. Sigma values in each panel represent the sum of the differences in normalized amplitude across all masker electrodes.

(three levels for each electrode), two blocks, and two conditions (each electrode in the first and second interval). Percent correct was calculated as the number of times (out of 36) that the more basal electrode in the pair was judged as higher in pitch. Percent-correct scores were converted to z scores so that linear regression analysis could be used to compare the ECAP separation index to pitch ranking.

III. RESULTS AND DISCUSSION

Table I lists the correlation coefficients and p values for the comparison between pitch-ranking z score and ECAP separation index for all electrode pairs per subject. Significant positive correlations were obtained for nine of the ten subjects (exception: M50). Correlation coefficients ranged

TABLE I. Correlation coefficients (r value) and significance (p value) for pitch-ranking z score vs ECAP separation index for all electrodes within each subject. Number of electrode pairs tested for electrode separations of 1, 2, 4, 5, and 6 for each subject.

			No. pairs at electrode separation of				
Subject	r	p	1	2	4	5	6
M15	0.60	0.001	8	8	6	⋯	4
M24	0.55	<0.001	12	11	9	⋯	7
M35b	0.81	<0.001	9	8	8	⋯	5
M50	0.59	0.29	1	1	1	1	1
M54b	0.68	0.005	3	7	3	⋯	2
R9	0.60	0.002	6	6	6	⋯	5
R13	1.00	0.002	⋯	2	⋯	⋯	1
R15	0.58	0.001	8	9	6	⋯	6
R21	0.52	0.02	5	5	5	⋯	5
R22	0.81	<0.001	6	5	3	⋯	3

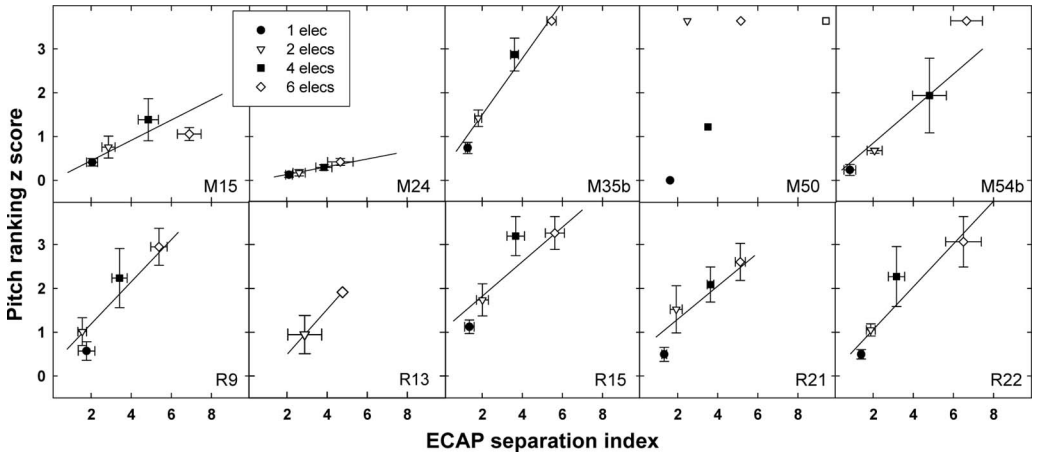


FIG. 2. Data for individual subjects showing mean (±1 SEM) z score as a function of mean (±1 SEM) ECAP separation index for all electrode pairs that were separated by 1, 2, 4, and 6 electrodes (separation of 5 electrodes for M50 is indicated by an open square). Solid lines represent linear regression analyses.

from 0.52 (p=0.02) to 1.0 (p=0.002)[3]. In contrast to the original study, these results suggest that the amount of separation between ECAP SOE functions is well correlated with electrode pitch-ranking accuracy.

Figure 2 shows pitch-ranking z score versus ECAP separation index plotted with the electrode-pair spacing as the parameter. Bidirectional error bars represent 1 standard error around the mean (±1 SEM). Different symbols represent different electrode separations, where a spacing of one is adjacent. All but two subjects (R13 and M50) had data for electrode separations of 1, 2, 4, and 6 electrodes. Table I lists the number of electrode pairs at each separation for individual subjects in Fig. 2. M50, M15, and R9 demonstrated nonmonotonic functions; however, the difference in z scores for electrode separations of 4 versus 6 for M15 and the difference in ECAP separation indices for separations of 1 versus 2 for R9 were not statistically significant (t test, $p>0.6$). In general, most subjects showed relatively linear functions, with larger z scores and larger ECAP separation indices for greater electrode separations.

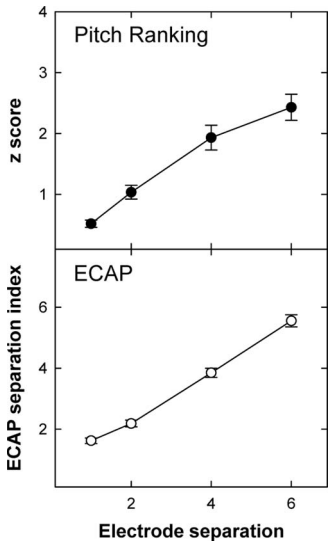


FIG. 3. Mean (±1 SEM) z scores (top panel) and mean (±1 SEM) ECAP separation indices (bottom panel) across all electrode pairs and subjects, plotted as a function of electrode separation.

Figure 3 shows average z scores (top) and ECAP separation indices (bottom) across all subjects for electrode separations of 1, 2, 4, and 6. There was a statistically significant correlation between mean z score and ECAP separation index as a function of electrode separation (r=0.98, p=0.019). Mean z scores and ECAP separation indices were significantly different across all electrode separations ($p<0.001$, one-way repeated-measures analysis of variance). In summary, pitch ranking improved significantly and ECAP spatial excitation patterns demonstrated significantly less overlap for increased spatial separation between electrode pairs.

IV. CONCLUSIONS

In contrast to the original study (Hughes and Abbas, 2006), a significant positive correlation between ECAP spatial excitation and electrode pitch ranking was found for nine of the ten subjects. In Hughes and Abbas (2006), pitch ranking was quantified as the slope of the psychometric function, in which the pitch of a fixed reference electrode was compared with the pitch of various other electrodes at progressively greater electrode separations. Therefore, the pitch-ranking measures reflected effects of electrode separation. In contrast, ECAP measures in that study were quantified as the width of the ECAP spatial excitation function at 75% of the normalized amplitude. The primary limitation of quantifying the ECAP data in this way is that the shape of the function at greater electrode separations is not characterized. For example, two ECAP functions may have similar widths at the 75% point, but they may have very different shapes overall [see Fig. 7 from Hughes and Abbas (2006)]. The present method takes into account the relative location of the edges of each ECAP pattern and the overall spread of each function. As a result, less overlap of ECAP spatial excitation patterns was found to be strongly correlated with greater accuracy of electrode pitch ranking. Future studies will evaluate whether the present methodology is sensitive enough to use ECAP SOE measures to predict pitch ranking with intermediate or virtual channels.

ACKNOWLEDGMENTS

Original data collection supported by the NIH, NIDCD, P50 DC00242; and NIH, NCRR, RR00059 (University of Iowa). New data analysis was supported by the NIH, NIDCD, R03 007017 (Boys Town National Research Hospital). Paul Abbas, Donna Neff, Walt Jesteadt, and two anonymous reviewers provided helpful editorial comments.

[1]Exception: M24 and M50, whose ECAP measures were made at 90% of the behavioral dynamic range. Psychophysical levels for those two subjects were 80%, 90%, and 100% of the dynamic range.

[2]In the psychophysical portion of the original study, each test electrode always occurred first within a block. Test electrodes were not necessarily always used as reference electrodes (i.e., second in a block).

[3]Because the regression line for R13 is based on only three data points, it yields little information about the goodness of fit. However, the slope of the line serves as a comparison to that of the other subjects.

Abbas, P. J., Hughes, M. L., Brown, C. J., Miller, C. A., and South, H. (**2004**). "Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential," Audiol. Neuro-Otol. **9**, 203–213.

Cohen, L. T., Richardson, L. M., Saunders, E., and Cowan, R. S. C. (**2003**). "Spatial spread of neural excitation in cochlear implant recipients: Comparison of improved ECAP method and psychophysical forward masking," Hear. Res. **179**, 72–87.

Eisen, M. D., and Franck, K. H. (**2005**). "Electrode interaction in pediatric cochlear implant subjects," J. Assoc. Res. Otolaryngol. **6**, 160–170.

Hughes, M. L., and Abbas, P. J. (**2006**). "The relation between electrophysiologic channel interaction and electrode pitch ranking in cochlear implant recipients," J. Acoust. Soc. Am. **119**, 1527–1537.

McKay, C. M., O'Brien, A., and James, C. J. (**1999**). "Effect of current level on electrode discrimination in electrical stimulation," Hear. Res. **126**, 159–164.

Acoustical scattering by radially stratified scatterers

Liang-Wu Cai
Department of Mechanical and Nuclear Engineering, Kansas State University, Manhattan, Kansas 66506

José Sánchez-Dehesa
Wave Phenomena Group, Department of Electronic Engineering, Polytechnic University of Valencia, C/Camino de Vera s.n., E-46022 Valencia, Spain

(Received 24 August 2007; revised 26 March 2008; accepted 2 June 2008)

A recursive solution procedure is developed to analyze the acoustical scattering by multilayer concentric circular cylindrical scatterers. The procedure is based on multiple scattering in the single-scatterer methodology originally proposed by one of the authors. The solution for a scatterer having an arbitrary number of layers is solved by recursively using the solution for a dual-layer scatterer. The solution procedure and its implementation are extensively validated. Computational characteristics, including potential pitfalls and their avoidance, are explored. The solution procedure can be used to analyze scatterers with continuously varying properties in the radial direction by approximating such a scatterer as a multilayer system. A comprehensive numerical example shows an interesting phenomenon: when a planar wave impinges onto a cooling tube in air, instead of forming a shadow zone, a strongly collimated beam is formed behind the cooling tube. This phenomenon can be explained as due to the gradient-index acoustical lens that is created around the cooling tube, which bends the sound and collimates it in the forward direction.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2967825]

PACS number(s): 43.20.Fn, 43.28.Bj, 43.35.Ty [LLT] Pages: 2715–2726

I. INTRODUCTION

Scatterers with radially varying physical properties are a common type of objects in the study of scattering of various types of waves. For example, the air surrounding a heated or cooled cylinder is radially stratified due to the temperature dependences of both the sound speed in the air and the mass density of the air. Polymeric optical fibers used in communication networks often exhibit radial gradients in their optical refractive index. Furthermore, multilayered and radially stratified cylinders are essential theoretical models that are often used as the first step toward modeling more complicated structures. For example, in medical ultrasonic imaging, some lesion areas do not possess uniform properties, and modeling such a region as a circular or a spherical scatterer with radially varying properties is the first step toward establishing a more realistic model.

On the other hand, the authors' interest in such scatterers stems from recent advances in metamaterials research. Metamaterials are man-made materials that exhibit certain unique or novel properties that are not available in constituent materials. The concept of metamaterials makes it possible to design scatterers with novel characteristics. A recent example is the gradient-index acoustic lens proposed by Torrent and Sánchez-Dehesa (2007). Another example is the recent demonstration of the cloaking of microwave by using such materials (Schurig *et al.*, 2006). A design of acoustic cloaking has been proposed based on the same design (Cummer and Schurig, 2007). One of the key ingredients in the Cummer-Schurig cloaking is the radially varying material properties.

Exact analytical solutions for scattering by radially stratified circular cylinders are scarce and are limited to extremely simple radial profiles. The most common approach to analyzing such a scatterer is to approximate it as consisting multiple homogeneous layers. When the layers become infinitesimally thin, it is expected that such a multilayer scatterer would approach the original scatterer. Scattering by multilayer scatterers has been studied rather extensively in the literature, which heavily inclines toward the scattering of electromagnetic waves. A significant portion of these efforts was focused on dual-layer concentric and eccentric scatterers and was reviewed by Cai (2004a,b). For multilayer scatterers, the analyses can be broadly categorized into two approaches.

The first approach is to simply list the continuity conditions at all interfaces and then solve the resulting coupled equation system. A recursive solution was developed by Bussey and Richmond (1975). However, the theoretical development of this work was only published as an internal report. Using this approach, Rao and Hamid (1975) analyzed the scattering of electromagnetic waves by multilayer dielectric cylinders, Ragheb and Hamid (1986) analyzed the multiple scattering by such cylinders, and Barabás (1987) studied the scattering of an oblique incident wave by such a cylinder. For elliptical multilayer cylinders, Kim and Yeh (1991) calculated the lossy cylinders, Kishk *et al.* (1992) studied eccentrically layered cylinders, Yousif and Elsherbeni (1999) analyzed such a cylinder subjected to an oblique incident wave, and Yin *et al.* (2000) studied their multiple scattering. Chati *et al.* (2005) used this approach to study the effects of a water layer among other solid layers in underwater acoustics. Ioannidou *et al.* (2004) analyzed the scattering of electromagnetic waves by eccentric multilayer cylinders that encloses multiple eccentric multilayer cylinders. This approach typically results in a large system of equations for a scatterer

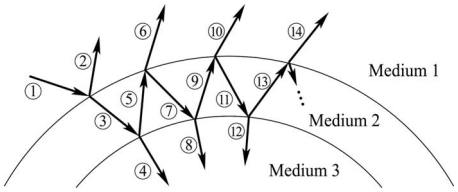


FIG. 1. Schematics for multiple scattering in a single scatterer.

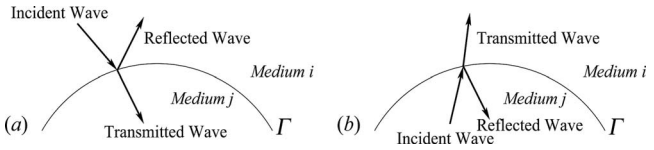


FIG. 2. Canonical problems: (a) first canonical problem and (b) second canonical problem.

with a large number of layers. Such a large equation system also risks becoming ill conditioned as the layers become very thin.

The second approach is often called the transfer-matrix method. The method was originally developed for analyzing elastic wave propagation in multilayered media (Thomson, 1950) and it treats a layer mathematically as a transformation between waves at the two surfaces of the layer. Using this approach, Yeh and Lindgren (1977) studied the scattering by a multilayer cylinder irradiated by a normal incident electromagnetic wave. Chew (1990) developed a recursive procedure for the scattering of electromagnetic waves by multilayer scatterers. Kluskens and Newman (1991) studied the multilayer chiral cylinder. Skelton and James (1993) combined this approach with the finite element formulation to study the propagation of acoustic waves in an anisotropic layered cylinder. Huang *et al.* (1995) analyzed the interphase effects on elastic wave propagation in fiber-reinforced composite materials in which the fibers are multilayered. Chiu and Chen (1995) studied the propagation of electromagnetic waves in a similar composite material. Helaly *et al.* (1997) studied a nonuniform plasma cylinder. Sastry and Munjal (1998a,b) studied the scattering of acoustic wave by a multilayer elastic cylinder embedded in a fluid medium. Caorsi and Pastorino (2004) used this approach to study multilayer isorefractive elliptic cylinders. Similarly, this approach also encounters computational difficulties when the layers become very thin. Kai and D'Alessio (1995) and Jiang *et al.* (2006) developed remedies by using alternative expressions for ratios of Bessel and Hankel functions to improve the computational range.

In this paper, a different recursive procedure is proposed. This approach is based on the concept of *multiple scattering in a single scatterer* developed by Cai (2004a,b). The concept has been used to analyze the scattering of elastic antiplane shear waves by dual-layer concentric (Cai, 2004b) and eccentric scatterers (Cai, 2005), and it has been validated to give analytically exact solutions. The advantage of this procedure is that it follows the physical process, and thus the solution maintains the same form for all types of waves. The solution for a dual-layer scatterer is used in this paper as the basis to analyze acoustical scattering by a scatterer having an arbitrary number of layers. The adaption for the eccentrically layered case can be obtained in a straightforward extension of the present solution procedure. The only complication is to ensure that layers are not entangled to produce geometrically unrealistic profiles.

This paper is organized as the follows: Section I gives a background overview. Section II reviews the solution for a dual-layer scatterer. Section III derives the recursive procedure for a general multilayer scatterer. Section IV validates the solution procedure using simple numerical examples. Section V discusses its computational characteristics. Section VI presents a comprehensive numerical example. Section VII presents the conclusions of the paper.

II. SOLUTION FOR A DUAL-LAYER SCATTERER

The idea of multiple scattering in a single scatterer recognizes that multiple scattering in its bare essence is further scattering of a scattered wave. This could occur in a layered scatterer, as depicted in Fig. 1. For instance, wave 5 is a wave scattered by the inner interface. It is then further scattered by the outer interface, forming waves 6 and 7; the resulting wave, 7, is further scattered by the inner interface, forming waves 8 and 9, and so on.

A. Two canonical problems

The solution for acoustical scattering by a dual-layer scatterer is obtained with the aid of two canonical problems, as depicted in Fig. 2. Each canonical problem is concerned with the interaction between an incident wave and a closed interface Γ that separates two media. The first canonical problem assumes an inward-traveling incident wave [Fig. 2(a)]; and the second canonical problem assumes an outward-traveling incident wave [Fig. 2(b)].

Without losing generality, denote the medium exterior of Γ as medium i and the medium enclosed by Γ as medium j. ϕ denotes the complex amplitude of the acoustic pressure. Using a polar coordinate system (r,θ), whose origin is located inside Γ, general expressions for various waves in both canonical problems are summarized in Table I. In Table I, all wave expansion coefficient matrices and wave expansion basis matrices are column matrices whose indices run from $-\infty$ to ∞. The expression for a specific wave can be written as the inner product of the coefficient matrix and the corresponding wave expansion basis matrix. For instance, the reflected wave in the second canonical problem is $\phi^R=\{\boldsymbol{B}\}^T\{\boldsymbol{J}_j(r,\theta)\}$. Such a matrix product represents the modal summation. Matrices $\{\boldsymbol{J}_j(r,\theta)\}$ and $\{\boldsymbol{H}_j(r,\theta)\}$ are the regular and singular wave expansion bases, respectively, in

TABLE I. Wave expansion coefficient and basis matrices for two canonical problems.

Wave	Symbol	Wave expansion coefficient	Wave expansion basis	
			First canonical	Second canonical
Incident	ϕ^I	$\{\boldsymbol{A}\}$	$\{\boldsymbol{J}_i(r,\theta)\}$	$\{\boldsymbol{H}_j(r,\theta)\}$
Reflected	ϕ^R	$\{\boldsymbol{B}\}$	$\{\boldsymbol{H}_i(r,\theta)\}$	$\{\boldsymbol{J}_j(r,\theta)\}$
Transmitted	θ^T	$\{\boldsymbol{C}\}$	$\{\boldsymbol{J}_j(r,\theta)\}$	$\{\boldsymbol{H}_i(r,\theta)\}$

medium j, whose entries at the nth row are $J_n(k_j r)e^{\hat{\imath} n\theta}$ and $H_n^{(1)}(k_j r)e^{\hat{\imath} n\theta}$, respectively, where $J_n(\cdot)$ and $H_n^{(1)}(\cdot)$ are Bessel and Hankel functions of the first kind at order n, respectively; k_j is the wave number in medium j, and $\hat{\imath}=\sqrt{-1}$.

For linear acoustics, wave expansion coefficient matrices of the reflected and the transmitted waves can be linearly related to that of the incident wave such that

$$\{\boldsymbol{B}\} = [\boldsymbol{\mathcal{R}}_{ij}]\{\boldsymbol{A}\}, \tag{1}$$

$$\{\boldsymbol{C}\} = [\boldsymbol{\mathcal{T}}_{ij}]\{\boldsymbol{A}\}, \tag{2}$$

where subscripts ij signify that the incident wave propagates from medium i to medium j. In this paper, the following conventions are adopted: layers are numbered inward, and the characteristic matrices for canonical problems are denoted by script-style typeface. Then, a matrix such as $[\boldsymbol{\mathcal{R}}_{ij}]$ also signifies the underlying canonical problems: the first canonical problem is when $i<j$, and the second canonical problem is when $i>j$.

Solutions to these canonical problems are obtained in processes similar to conventional single-scatterer problems: by enforcing the continuities of the pressure and the normal component of the particle velocity at the interface Γ. For a circular interface of radius a, with the origin located at the center of Γ, these boundary conditions can be written as

$$\phi^T|_{r=a} = (\phi^I + \phi^R)|_{r=a}, \tag{3}$$

$$\frac{1}{\rho_j}\left.\frac{\partial \phi^T}{\partial r}\right|_{r=a} = \frac{1}{\rho_i}\left.\left(\frac{\partial \phi^I}{\partial r} + \frac{\partial \phi^R}{\partial r}\right)\right|_{r=a}. \tag{4}$$

Consequently, the solutions to the two canonical problems for acoustic media are, assuming $i<j$,

$$[\mathcal{R}_{ij}]_{nn} = -\frac{z_i J_n'(k_j a)J_n(k_i a) - z_j J_n(k_j a)J_n'(k_i a)}{z_i J_n'(k_j a)H_n^{(1)}(k_i a) - z_j J_n(k_j a)H_n^{(1)\prime}(k_i a)}, \tag{5}$$

$$[\mathcal{T}_{ij}]_{nn} = -\frac{2\hat{\imath}}{\pi k_i a}\frac{z_j}{z_i J_n'(k_j a)H_n^{(1)}(k_i a) - z_j k J_n(k_j R)H_n^{(1)\prime}(k_i a)}, \tag{6}$$

$$[\mathcal{R}_{ji}]_{nn} = -\frac{z_i H_n^{(1)\prime}(k_j a)H_n^{(1)}(k_i a) - z_j H_n^{(1)}(k_j a)H_n^{(1)\prime}(k_i a)}{z_i J_n'(k_j a)H_n^{(1)}(k_i a) - z_j J_n(k_j a)H_n^{(1)\prime}(k_i a)}, \tag{7}$$

$$[\mathcal{T}_{ji}]_{nn} = -\frac{2\hat{\imath}}{\pi k_j a}\frac{z_i}{z_i J_n'(k_j a)H_n^{(1)}(k_i a) - z_j J_n(k_j a)H_n^{(1)\prime}(k_i a)}, \tag{8}$$

where $z_i=\rho_i c_i$ is the specific acoustic impedance for medium i.

B. Solution for a dual-layer scatterer

The solution for the scattering of antiplane elastic shear waves by a dual-layer concentric scatterer has been obtained by Cai (2004a,b). The solution follows the physical process of multiple scattering in a single scatterer. Thus the solution maintains its form for all types of waves, with the only difference being the solutions to the associated canonical problems. This solution is adapted for acoustical scattering and is summarized below.

Denote the host as medium 1, the intermediate layer as medium 2, and the core of the scatterer as medium 3, as shown in Fig. 1. Introduce the following notations for various wave fields:

$$\phi^{\text{inc}} = \{\boldsymbol{A}\}^T\{\boldsymbol{J}_1(r,\theta)\}, \tag{9}$$

$$\phi^{\text{scr}} = \{\boldsymbol{B}\}^T\{\boldsymbol{H}_1(r,\theta)\}, \tag{10}$$

$$\phi^{\text{ref}} = \{\boldsymbol{C}\}^T\{\boldsymbol{J}_3(r,\theta)\}, \tag{11}$$

$$\phi^{\text{inter}} = \{\boldsymbol{D}\}^T\{\boldsymbol{J}_2(r,\theta)\} + \{\boldsymbol{E}\}^T\{\boldsymbol{H}_2(r,\theta)\}, \tag{12}$$

where superscripts inc, scr, ref, and inter signify the incident wave, the scattered wave, the refracted wave, and the wave in the intermediate layer, respectively. Furthermore, denote the linear transformations between the resulting waves and the incident wave as

$$\{\boldsymbol{B}\} = [\boldsymbol{R}]\{\boldsymbol{A}\}, \tag{13}$$

$$\{\boldsymbol{C}\} = [\boldsymbol{T}]\{\boldsymbol{A}\}, \tag{14}$$

$$\{\boldsymbol{D}\} = [\boldsymbol{S}]\{\boldsymbol{A}\}, \tag{15}$$

$$\{\boldsymbol{E}\} = [\boldsymbol{F}]\{\boldsymbol{A}\}. \tag{16}$$

Then, these characteristic matrices are found to be (Cai, 2004a,b)

$$[\boldsymbol{S}] = ([\boldsymbol{I}] - [\boldsymbol{\mathcal{R}}_{21}][\boldsymbol{\mathcal{R}}_{23}])^{-1}[\boldsymbol{\mathcal{T}}_{12}], \tag{17}$$

$$[\boldsymbol{R}] = [\boldsymbol{\mathcal{R}}_{12}] + [\boldsymbol{\mathcal{T}}_{21}][\boldsymbol{\mathcal{R}}_{23}][\boldsymbol{S}], \tag{18}$$

$$[\boldsymbol{F}] = [\boldsymbol{\mathcal{R}}_{23}][\boldsymbol{S}], \tag{19}$$

$$[\boldsymbol{T}] = [\boldsymbol{\mathcal{T}}_{23}][\boldsymbol{S}]. \tag{20}$$

III. SOLUTION FOR A GENERAL MULTILAYER SCATTERER

In this section, acoustical scattering by a general concentric multilayer circular cylindrical scatterer is considered. The word "general" means that the number of layers is arbitrary. Assume that all layers are homogeneous.

A. Problem setup

The layer structure for the multilayer scatterer is depicted in Fig. 3. The scatterer consists of N layers, radii r_i $(i=1,2,\ldots,N)$, and $r_1>r_2>\cdots>r_N$. A layer bounded by r_{i-1} and r_i is referred to as layer i. The host, occupying the region $r>r_1$, is referred to as layer 0 when necessary. However, for notational simplicity, field quantities for this layer would not carry a subscript. The innermost layer may be referred to as the *core*.

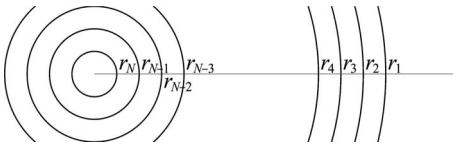


FIG. 3. Multilayered structure.

In the host medium, there exist both the incident and the scattered waves, which are expressible as

$$\phi^{\text{inc}} = \{\boldsymbol{A}\}^T\{\boldsymbol{J}(k,\boldsymbol{r})\}, \tag{21}$$

$$\phi^{\text{scr}} = \{\boldsymbol{B}\}^T\{\boldsymbol{H}(r,\theta)\}. \tag{22}$$

A transmitted wave in the innermost layer exists only when it is penetrable, in which case, it is expressible as

$$\phi^{\text{trs}} = \{\boldsymbol{C}\}^T\{\boldsymbol{J}_N(r,\theta)\}. \tag{23}$$

In all other layers, for layer $i(1 \leqslant i \leqslant N-1)$, the waves are expressible as

$$\phi^i = \{\boldsymbol{D}_i\}^T\{\boldsymbol{J}_i(r,\theta)\} + \{\boldsymbol{E}_i\}^T\{\boldsymbol{H}_i(r,\theta)\}. \tag{24}$$

Similar to the dual-layer case, introduce the following characteristic matrices that relate the wave expansion coefficient matrices of the resulting waves to that of the incident wave

$$\{\boldsymbol{B}\} = [\boldsymbol{R}]\{\boldsymbol{A}\}, \tag{25}$$

$$\{\boldsymbol{C}\} = [\boldsymbol{T}]\{\boldsymbol{A}\}, \tag{26}$$

$$\{\boldsymbol{D}_i\} = [\boldsymbol{S}_i]\{\boldsymbol{A}\}, \tag{27}$$

$$\{\boldsymbol{E}_i\} = [\boldsymbol{F}_i]\{\boldsymbol{A}\}. \tag{28}$$

The task at hand is to determine the expressions for these characteristic matrices. The solution strategy is to repeatedly use the dual-layer solution process described in Sec. II B. In the discussions that follow, layers in the dual-layer case (depicted in Fig. 1) are referred to as *medium X*, and layers in the multilayer case (depicted in Fig. 3) are referred to as *layer X*.

B. The first step

In the first step, the innermost two layers are analyzed: layer N is treated as medium 3; layer $N-1$ is treated as medium 2; and layer $N-2$ is treated as medium 1. Directly adopting the solution for the dual-layer case, this step gives

$$[\boldsymbol{S}_{N-1}] = ([\boldsymbol{I}] - [\boldsymbol{\mathcal{R}}_{(N-1)(N-2)}][\boldsymbol{\mathcal{R}}_{(N-1)N}])^{-1}[\boldsymbol{\mathcal{T}}_{(N-2)(N-1)}], \tag{29}$$

$$[\boldsymbol{R}_{N-1}] = [\boldsymbol{\mathcal{R}}_{(N-2)(N-1)}] + [\boldsymbol{\mathcal{T}}_{(N-1)(N-2)}][\boldsymbol{\mathcal{R}}_{(N-1)N}][\boldsymbol{S}_{N-1}], \tag{30}$$

$$[\boldsymbol{F}_{N-1}] = [\boldsymbol{\mathcal{R}}_{(N-1)N}][\boldsymbol{S}_{N-1}], \tag{31}$$

$$[\boldsymbol{T}] = [\boldsymbol{\mathcal{T}}_{(N-1)N}][\boldsymbol{S}_{N-1}]. \tag{32}$$

C. The second step

In this step, the innermost two layers (layers N and $N-1$) are treated as a composite core (medium 3), layer $N-2$ becomes medium 2, and layer $N-3$ becomes medium 1. There are two significant differences from the previous step. First, matrix $[\boldsymbol{\mathcal{R}}_{23}]$ in the dual-layer case represents the reflection from the core. With the composite core, this matrix should be replaced by the matrix that represents the total reflection into the host in the previous step, that is, $[\boldsymbol{R}_{N-1}]$. Second, all the waves in layers N and $N-1$ will be modified from the previous step since the wave impinging into layer $N-1$ is no longer the incident wave.

Introduce an intermediate series of matrices $[\boldsymbol{L}_i]$ that represents the incident wave into medium 3 in an intermediate step when layer i is treated as medium 2. This way, $[\boldsymbol{S}_i]$ maintains its meaning as representing the standing-wave component in layer i, while $[\boldsymbol{L}_i]$ represents the layering effect. Thus, in this step,

$$[\boldsymbol{L}_{N-2}] = ([\boldsymbol{I}] - [\boldsymbol{\mathcal{R}}_{(N-2)(N-3)}][\boldsymbol{R}_{N-1}])^{-1}[\boldsymbol{\mathcal{T}}_{(N-3)(N-2)}]. \tag{33}$$

Back-tracking one step, $[\boldsymbol{L}_{N-1}]$ would be identical to $[\boldsymbol{S}_{N-1}]$ in the previous step. Then, according to the solution for the dual-layer case, the waves produced by this multilayer scatterer in media 1 and 2 are the following:

$$[\boldsymbol{S}_{N-2}] = [\boldsymbol{L}_{N-2}], \tag{34}$$

$$[\boldsymbol{R}_{N-2}] = [\boldsymbol{\mathcal{R}}_{(N-3)(N-2)}] + [\boldsymbol{\mathcal{T}}_{(N-2)(N-3)}][\boldsymbol{R}_{N-1}][\boldsymbol{L}_{N-2}], \tag{35}$$

$$[\boldsymbol{F}_{N-2}] = [\boldsymbol{R}_{N-1}][\boldsymbol{L}_{N-2}]. \tag{36}$$

Within the composite core, as noted earlier, it is the incoming wave in layer $N-2$, represented by $[\boldsymbol{L}_{N-2}]$, that acts as the incident wave. Accordingly, the expressions for the characteristic matrices for all the waves in the composite core are modified by right-multiplying by $[\boldsymbol{L}_{N-2}]$; that is,

$$[\boldsymbol{S}_{N-1}] = [\boldsymbol{L}_{N-1}][\boldsymbol{L}_{N-2}], \tag{37}$$

$$[\boldsymbol{F}_{N-1}] = [\boldsymbol{\mathcal{R}}_{(N-1)N}][\boldsymbol{L}_{N-1}][\boldsymbol{L}_{N-2}], \tag{38}$$

$$[\boldsymbol{T}] = [\boldsymbol{\mathcal{T}}_{(N-1)N}][\boldsymbol{L}_{N-1}][\boldsymbol{L}_{N-2}]. \tag{39}$$

D. Recursive procedure

The above procedure can be used recursively until medium 1 becomes the actual host medium. This process can be summarized as the following two coupled recursive equations:

$$[\boldsymbol{L}_i] = ([\boldsymbol{I}] - [\boldsymbol{\mathcal{R}}_{i(i-1)}][\boldsymbol{R}_{i+1}])^{-1}[\boldsymbol{\mathcal{T}}_{(i-1)i}], \tag{40}$$

$$[\boldsymbol{R}_i] = [\boldsymbol{\mathcal{R}}_{(i-1)i}] + [\boldsymbol{\mathcal{T}}_{i(i-1)}][\boldsymbol{R}_{i+1}][\boldsymbol{L}_i]. \tag{41}$$

The recursion starts from $i=N-1$ and $[\boldsymbol{R}_N]=[\boldsymbol{\mathcal{R}}_{(N-1)N}]$. The recursion moves outward, in the direction of decreasing i, until $i=1$. To the end, the total scattered wave is represented by

$$[\boldsymbol{R}] = [\boldsymbol{R}_1], \tag{42}$$

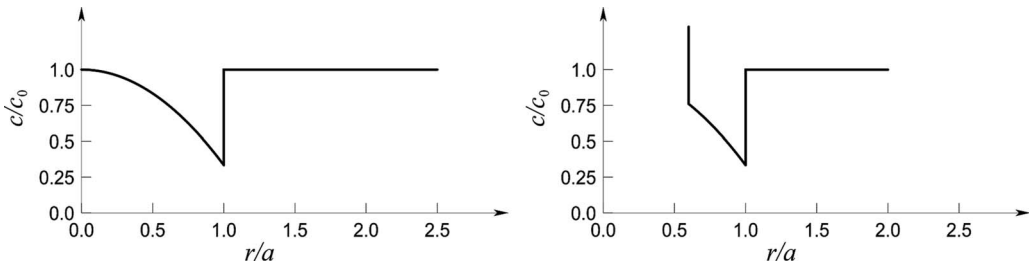


FIG. 4. Profiles of the sound speed for the two test cases, where c_0 is the sound speed of the host medium. Left: coreless case. Right: rigid-core case.

which is what is conventionally referred to as the *T-matrix* for the multilayer scatterer. Characteristic matrices for layer i can be calculated as follows:

$$[\boldsymbol{S}_i] = \prod_{j=1}^{i} [\boldsymbol{L}_j], \tag{43}$$

$$[\boldsymbol{F}_i] = [\boldsymbol{R}_{i+1}][\boldsymbol{S}_i], \tag{44}$$

where the cumulative matrix product is left-multiplied by matrices with larger j values, as illustrated in Eq. (37). Moreover, for the core,

$$[\boldsymbol{T}] = [\boldsymbol{\mathcal{T}}_{N(N+1)}][\boldsymbol{S}_{N-1}]. \tag{45}$$

It is interesting to note that the core of the above recursive procedure [Eqs. (40) and (41)] is formally the same as the recursive relation obtained by Chew [Eq. (3.2.19)] in Chew (1990).

IV. VALIDATIONS

A. A limiting case

As the first validation, consider a limiting case in which a uniform scatterer is divided into N layers. Since all intermediate layers are made of the same material, there is only total transmission and no scattering at these internal interfaces. The solutions to the canonical problems are $[\boldsymbol{\mathcal{R}}_{ij}]=0$ and $[\boldsymbol{\mathcal{T}}_{ij}]=[\boldsymbol{\mathcal{I}}]$ for all i and j except for the canonical problems at the scatterer's exterior boundary r_1, that is, when either $i=0$ or $j=0$.

The recursion using Eqs. (40) and (41) gives $[\boldsymbol{L}_i]=0$ and $[\boldsymbol{R}_i]=0$ for $i=2,3,\ldots,N-1$, and $[\boldsymbol{L}_1]=[\boldsymbol{\mathcal{T}}_{01}]$ and $[\boldsymbol{R}_1]=[\boldsymbol{\mathcal{R}}_{01}]$. Then, according to Eqs. (43) and (44), within layer i $(1\leqslant i\leqslant N-1)$, $[\boldsymbol{S}_i]=[\boldsymbol{\mathcal{T}}_{01}]$, $[\boldsymbol{F}_i]=0$, and in the core, $[\boldsymbol{T}]=[\boldsymbol{\mathcal{T}}_{01}]$. That is, in all layers, the resulting waves are identical to the transmitted wave in a single-scattering problem.

B. Two radially stratified scatterers

Two test cases are used for further validations. In both cases, the host is water, with a sound speed of 1500 m/s and a mass density of 1000 kg/m^3. In the first case, the scatterer has a constant mass density of 1500 kg/m^3, and the sound speed varies according to

$$c = 1500 - 1000(r/a)^2 \text{ m/s}, \tag{46}$$

where a is the outer radius of the scatterer. This case is referred to as the *coreless case*. In the second case, a rigid core of radius $0.6a$ is coated by a layer, within which the sound speed changes in accordance with Eq. (46). This case is referred to as the *rigid-core case*. The sound speed profiles for both cases are sketched in Fig. 4

In the computations, the scatterers are evenly divided into a predetermined number of layers, and the properties at the outer edge of a layer are taken as the layer's uniform properties. For the rigid-core case, the core is never subdivided. The infinite series of summations (the matrices of infinite sizes) are truncated at the N_tth term, where N_t is the smallest integer such that $|[R]_{nn}|<10^{-10}$ when $n>N_t$.

1. Property of T-matrix validation

A well-known property of T-matrix, the *unitarity* of the T-matrix [see, e.g., Waterman (1970)] is used as one of the validations. This property requires a T-matrix, which is $[\boldsymbol{R}]$ in the present paper, to satisfy the following relation:

$$-\Re\{[\boldsymbol{R}]\} = [\boldsymbol{R}]^*[\boldsymbol{R}], \tag{47}$$

where $*$ denotes the conjugate transposition of a matrix and $\Re$ denotes the real part of a complex quantity. This property is a consequence of energy conservation for a lossless system. Since $[\boldsymbol{R}]$ is diagonal for a concentric circular shell system, this property requires all entries of the $[\boldsymbol{R}]$ matrix to satisfy the following relation:

$$-\Re\{[R]_{nn}\} = |[R]_{nn}|^2. \tag{48}$$

Equation (48) implies that (1) the real part of $[R]_{nn}$ must be negative, (2) the magnitude of the real part of $[R]_{nn}$ must be smaller than 1, and (3) $|\Im\{[R]_{nn}\}| = \sqrt{-\Re\{[R]_{nn}\} - \Re\{[R]_{nn}\}^2}$; that is, the imaginary part can be calculated from the real part.

TABLE II. Values for $[R]_{00}$ entry for two test cases.

Number of layers	Coreless case	Rigid-core case
2	−0.937 986+0.241 181i	−0.675 890+0.468 041i
4	−0.595 440+0.490 807i	−0.246 839+0.431 172i
6	−0.478 576+0.499 541i	−0.167 774+0.373 665i
8	−0.421 851+0.493 855i	−0.137 202+0.344 060i
10	−0.388 634+0.487 440i	−0.121 291+0.326 465i
20	−0.324 535+0.468 201i	−0.094 220 4+0.292 135i
50	−0.287 848+0.452 760i	−0.080 639 0+0.272 280i
100	−0.275 955+0.446 994i	−0.076 505 1+0.265 805i
200	−0.270 076+0.443 999i	−0.074 507 5+0.262 595i
300	−0.268 126+0.442 984i	−0.073 851 6+0.261 529i
400	−0.267 153+0.442 473i	−0.073 525 6+0.260 997i

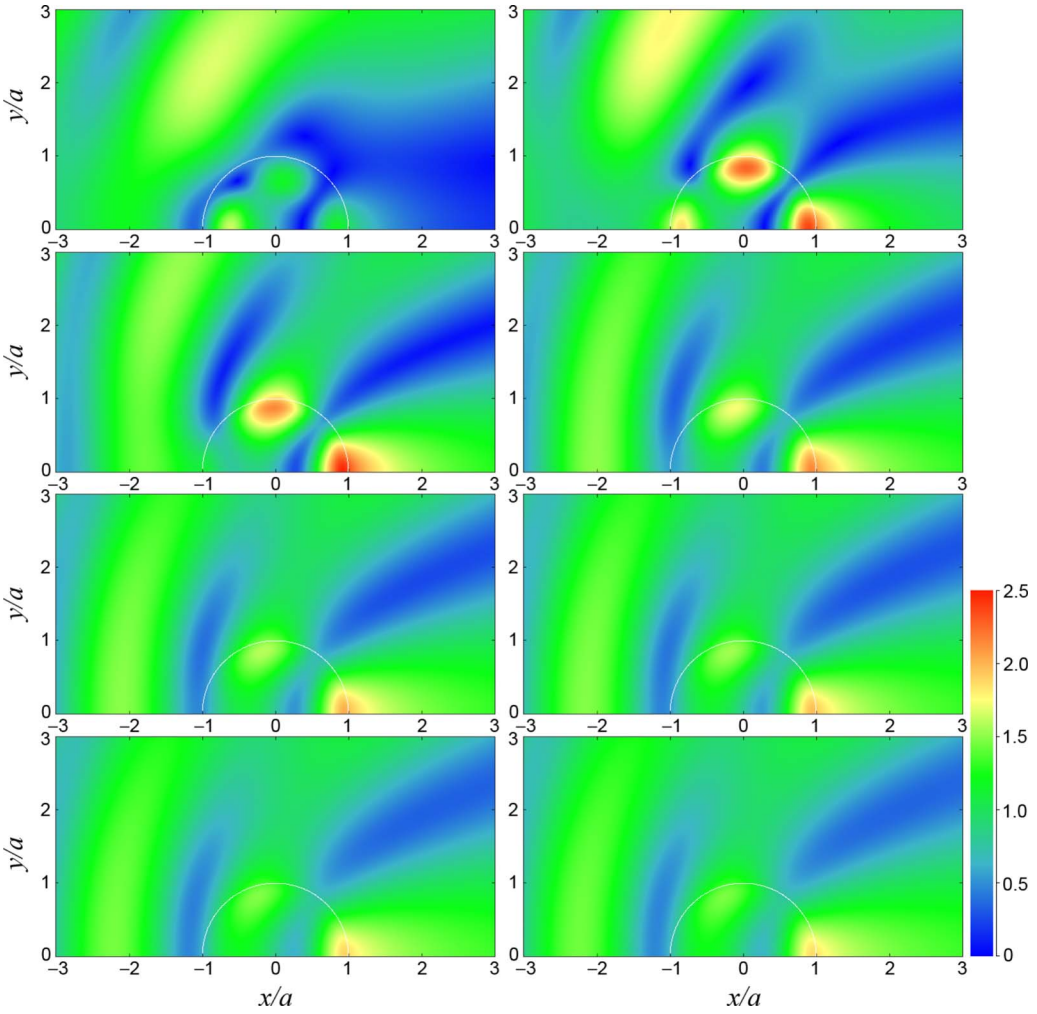


FIG. 5. (Color online) Distributions of amplitude of acoustic pressure for the coreless case at $ka=1.5$. The numbers of layers used in the computation are 2, 4 (top row), 6, 10 (second row), 15, 20 (third row), 50, and 100 (bottom row).

The calculated values of the $[R]_{00}$ entry at a normalized frequency $ka=1.5$ are listed in Table II when different numbers of layers are used in both cases. It can be readily verified that all the listed values satisfy the above relation.

2. Pressure continuity validations

The continuity of acoustic pressure is one of the conditions that define the canonical problems. This continuity must be satisfied at all interfaces, including the internal interfaces. Figures 5 and 6 show the pressure amplitude distributions in the vicinity of the scatterer for both cases, including the interior of the scatterer, at a normalized frequency $ka=1.5$. The pressure amplitude is given as $|\phi|$ normalized by that of the incident wave. In both figures, the numbers of layers used in the computation are 2 and 4 in the top row, 6 and 10 in the second row, 15 and 20 in the third row, and 50 and 100 in the bottom row. Due to symmetry, only the upper half space is shown for each case. The white circles delineate the exterior boundary of the scatterer and the rigid core. Pressure continuity is evident in both figures.

3. Convergence validation

It is expected that as the number of layers increases, the multilayer system approaches a scatterer whose properties vary continuously. Consequently, all related quantities, from pressure amplitude to characteristic matrices, approach their respective stable limits.

The values for $[R]_{00}$ entries listed in Table II for both cases show a clear trend of convergence as the number of layers increases. The pressure amplitude distributions for both cases in Figs. 5 and 6 also show a clear trend of approaching respective limiting pressure distributions. Furthermore, it can be observed that the field distribution does not show any discernible difference when 20, 50, and 100 layers are used in the computations. This suggests that in this case, using 20 layers produces generally satisfactory approximations.

Based on the above validations, it can be concluded that both the solution procedure and its implementation are correct. Furthermore, for layered structures, the solution thus obtained is analytically exact.

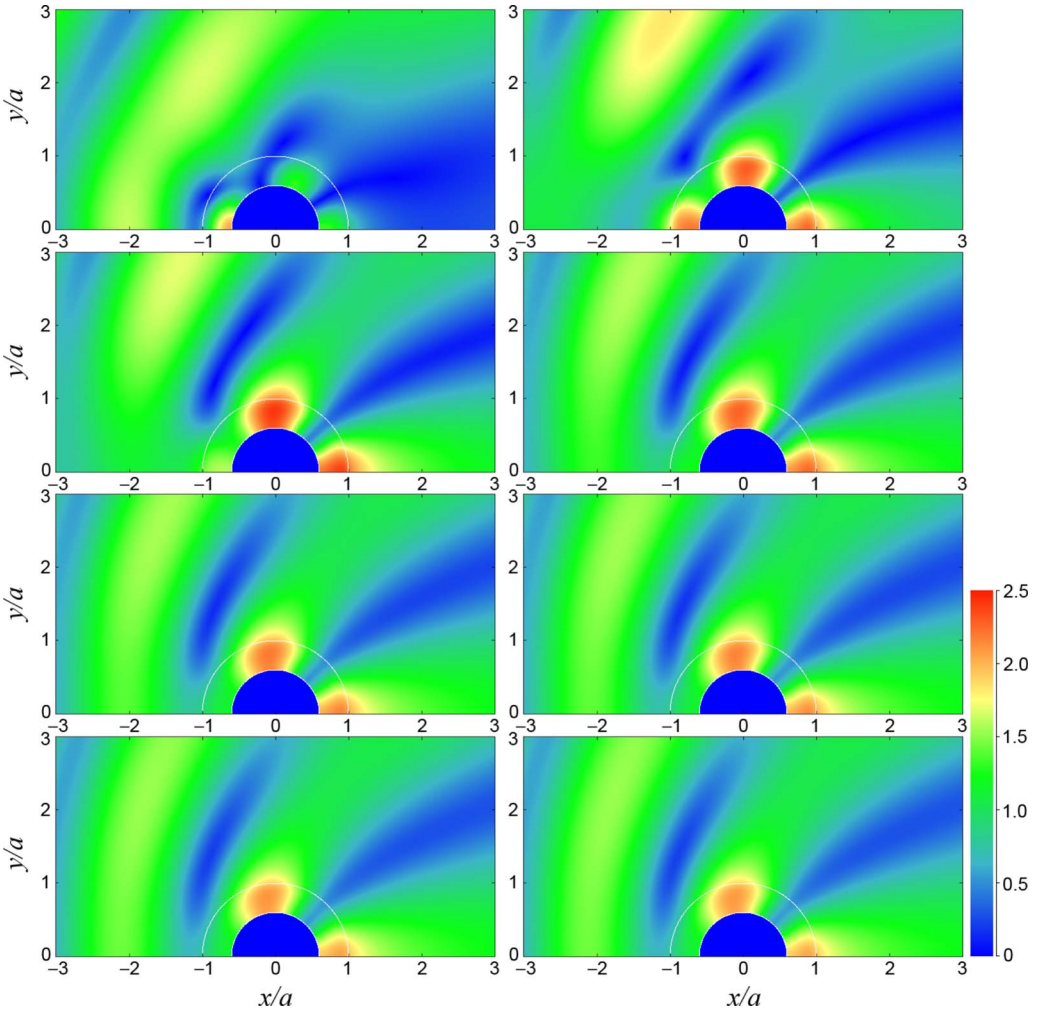


FIG. 6. (Color online) Distributions of amplitude of acoustic pressure for the rigid-core case at $ka=1.5$. The numbers of layers used in the computation are 2, 4 (top row), 6, 10 (second row), 15, 20 (third row), 50, and 100 (bottom row).

V. COMPUTATIONAL CHARACTERISTICS

The solution procedure described in Sec. III has the following characteristics:

(i) The scatterers under consideration consist of concentric circular interfaces. Thus, all characteristic matrices are diagonal.

(ii) Although the examples used in the validation use evenly divided layers, there is no such restriction imposed by the solution procedure.

(iii) Although Eq. (40) contains a matrix inversion, its computation requires little effort.

(iv) The solution is analytically exact for discretely layered scatterers. It can be used to approximate the scatterer with continuously varying properties.

(v) The current solution can be extended to eccentrically layered shell systems when the recursion is based on the solution for an eccentric dual-layer scatterer, whose solution has been obtained by Cai (2005). This extension will not be further elaborated, except to point out that in such cases, characteristic matrices become full square matrices, and matrix inversion is no longer a trivial matter. However, when the layers are sufficiently thin, approximation similar to $(1+x)^{-1}\approx 1+x$ can be adopted.

Two other important computational issues are discussed in detail in this section.

A. Numerical overflow and underflow

When the number of truncation terms is large, both numerical underflow and overflow could occur: for a large n, underflow would occur when calculating $J_n(x)$ and overflow would occur for $H_n^{(1)}(x)$. This is typical for most scattering analyses due to the properties of Bessel functions: when the order n is much greater than the argument x, $J_n(x)$ decreases exponentially and $H_n^{(1)}(x)$ increases exponentially, while $J_n(x)H_n^{(1)}(x)=O(1)$.

However, the over- and underflows may occur prematurely due to the numerators in Eqs. (5) and (8). Specifically, the two multiplications of $J_n(\cdot)$ and $J_n'(\cdot)$ in Eq. (5) may cause premature underflows, and the two multiplications of $H_n^{(1)}(\cdot)$ and $H_n^{(1)'}(\cdot)$ in Eq. (8) may cause premature overflows.

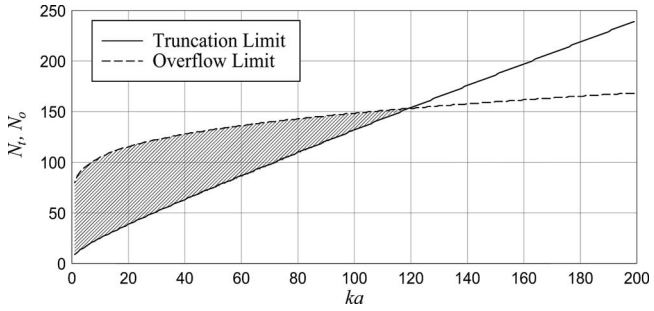


FIG. 7. Truncation limit (solid curve) as determined by Eq. (49) and the overflow limited (dashed curve) as determined by Eq. (50). The intersection point signifies the upper limit of the computable frequency. The shaded area defines the region where a computation with a desired accuracy is possible.

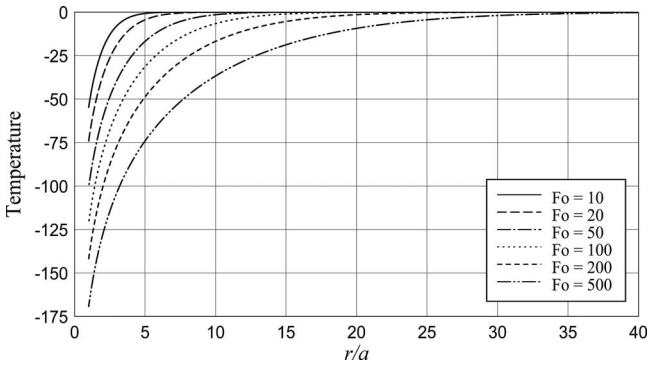


FIG. 8. Distributions of temperature in the vicinity of the cooling tube at six different Fo values: 10, 20, 50, 100, 200, and 500.

The following strategy can be used to circumvent such premature over- and underflows. Suppose that the smallest representable double precision floating number in a computer is in the order of 10^{-320}. The calculation of $J_n'(k_j a)J_n(k_i a)$ in Eq. (5), for example, is performed in the following three-step sequence: First, $J_n'(k_j a)$ is calculated. It is then scaled up by a large number M, say, 10^{300}, to give $J_n'(k_j a)M$. Finally, it is multiplied by $J_n(k_i a)$. A similar procedure is simultaneously adopted for calculating the numerator in Eq. (8), except that the scaling is downward, by dividing with M. This way, all $[\boldsymbol{\mathcal{R}}_{ij}]$ matrices are scaled up by a factor M, and $[\boldsymbol{\mathcal{R}}_{ji}]$ matrices are scaled down by a factor M. The scaled $[\boldsymbol{\mathcal{R}}_{ij}]$ and $[\boldsymbol{\mathcal{R}}_{ji}]$ matrices are used throughout the recursion of Eqs. (40) and (41). The resulting $[\boldsymbol{R}_i]$ matrices are scaled up by a factor M while $[\boldsymbol{L}_i]$ matrices remain correct. To obtain the correct result, all $[\boldsymbol{R}_i]$ matrices are scaled down by the factor M before computing other characteristics matrices.

B. Upper limit of frequency

There exists an upper limit of frequency beyond which a desired accuracy may not be achieved. This could become a major computational difficulty at high frequencies when the innermost radius is significantly smaller than the outermost radius, which is typical of coreless scatterers.

To illustrate the mechanism of this limit, consider an oversimplified example in which only the overflow of $H_n^{(1)}(\cdot)$ is of concern. Assume that the outermost and innermost radii of the scatterer are a and $0.01a$, respectively. To ensure a desired accuracy, the truncation term N_t should be sufficiently large such that $|H_n^{(1)}(ka)|$ is smaller than the desired error tolerance ε for any $n > N_t$. On the other hand, N_t must not exceed N_o, for which $H_n^{(1)}(0.01ka)$ could overflow for any $n > N_o$.

The variations of both N_t and N_o with the ka value are shown in Fig. 7. The truncation limit is the smallest N_t that satisfies

$$|H_{N_t}^{(1)}(ka)| \leqslant 10^{-6}. \tag{49}$$

The overflow limit is the smallest N_o that satisfies

$$|H_{N_o}^{(1)}(0.01ka)| \geqslant 10^{300}. \tag{50}$$

In order to perform computations with the desired accuracy, the truncation term should fall within the region above the curve labeled as the "truncation limit" and below the curve labeled as the "overflow limit," shown as a shaded region in Fig. 7. The intersection point of the two curves is the upper limit of ka beyond which computations with expected accuracy are not possible. At this moment, the authors have not found a good resolution to this problem, except to suggest the use of a larger innermost radius if the problem permits.

VI. NUMERICAL EXAMPLE

In this section, a comprehensive example is presented as an application of the present solution procedure to a radially stratified scatterer. This example offers a novel way to analyze scattering by a rigid scatterer in a nonuniform host medium and an interesting physical phenomenon.

A. Problem description

Assume that an infinitely long cooling tube runs through the air. Due to its cooling effect, the air temperature in its vicinity becomes radially stratified, which, in turn, causes both the sound speed and the mass density of the air to change. Assume that, in the vicinity of the tube, air temperature can be approximated by that due to an ideal line heat source of negative strength (ideal heat sink) as (Carslaw and Jaeger, 1959)

$$T(r,t) = -\frac{Q}{2\pi k_{\text{th}}}\text{Ei}\left(-\frac{r^2}{4\alpha t}\right), \tag{51}$$

where k_{th} and α are the thermal conductivity and thermal diffusivity, respectively, of the air, Q is the strength of the heat sink, t is the time, T is the temperature increase from the ambient value, which is assumed to be 0 °C in this study, and Ei$(\cdot)$ is the exponential integral function. Using *Fourier number* as a nondimensional time,

$$\text{Fo} = \frac{\alpha t}{4a^2}, \tag{52}$$

where a is the radius of the cooling tube. Denoting $T_0 = Q/(2\pi k_{\text{th}})$ and $\bar{r} = r/a$, then

$$T(r,t) = T_0 E_1\left(\frac{\bar{r}^2}{\text{Fo}}\right), \tag{53}$$

where $E_1(x) = -\text{Ei}(-x)$. Assuming $T_0 = -30$ °C, Fig. 8 shows

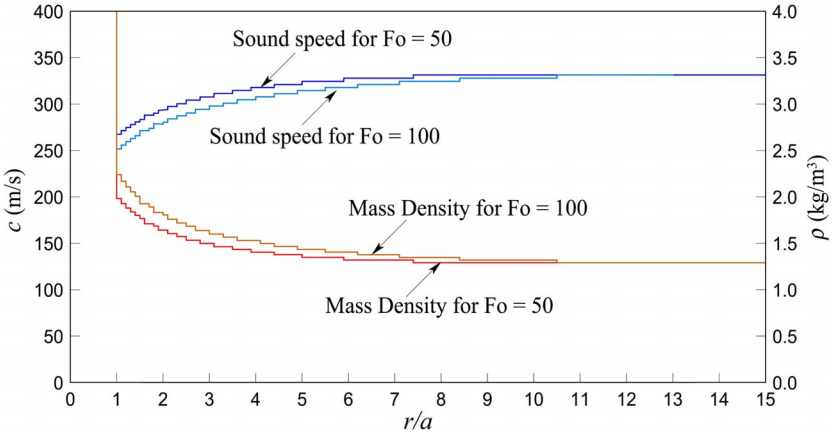


FIG. 9. (Color online) Scatterer profiles for Fo=50 and 100. Steps indicate the layering structure.

the distributions of temperature at six different Fo values: 10, 20, 50, 100, 200, and 500.

The sound speed in the air and the air's mass density vary with the temperature according to the following relations [see, e.g., Beranek (1986)]:

$$c = 331.3\sqrt{1 + \frac{T}{273.15}}, \tag{54}$$

$$\rho = \frac{p}{R(273.15 + T)}, \tag{55}$$

where T is the temperature in °C, p is the standard atmospheric pressure, which is 101.325 kPa at 0 °C, and R =287.05 J/(kg K).

The scatterer consists of the cooling tube plus a cylindrical annulus of air in its vicinity, bounded by the radius where the sound speed drops 1% from its ambient value of 331.3 m/s. The layer structure of the cylindrical shell is constructed using the following procedure. First, the air temperature is calculated according to Eq. (53) for a range of radius at an interval of $0.1a$. The sound speed and mass density are subsequently calculated for the same series of positions using Eqs. (54) and (55). Moving toward the core, the sound speed at each position is compared with the ambient value, and an interlayer interface is identified when the sound speed variation exceeds integer multiples of 1%. The sound speed and the mass density at a layer's outer edge is taken as the uniform sound speed and mass density for the layer. This way, the smallest thickness of a layer is $0.1a$.

Two Fourier numbers Fo=50 and 100 are arbitrarily chosen for analysis. For these two Fo numbers, the scatterer has an outermost radius of $7.4a$ and $10.5a$, respectively. It consists of 20 and 24 layers, respectively, and the temperature at the surface of the tube reaches −100.6 and −121.1 °C, respectively. The cross-sectional profiles of the sound speed and mass density of the scatterer are shown in Fig. 9. The discrete steps in Fig. 9 reflect the layer structures. The incident wave is a planar wave of unit amplitude traveling in the $+x$-direction.

B. Pressure distribution

Figure 10 shows the amplitudes of acoustic pressure in the vicinity of the cooling tube for Fo=50 and 100 at frequencies ka=2, 4, and 6, in comparison with those due to a rigid cylinder. As before, the pressure is given by $|\phi|$ normalized by that of the incident wave. The most prominent feature in Fig. 10 is several highly collimated beams formed in what originally is the shadow region. The main beam in the forward direction exceeds the strength of the incident wave. As the frequency increases, the number of beams increases while the beams become narrower. As the temperature deceases, the strength of these beams increases. This behavior has been consistently observed for the computed frequency range up to ka=10.

C. Form factor

A scattering form factor, which describes the angular distribution of the scattered wave in the far field, could provide more insight into this interesting phenomenon. The far-field approximation uses the following asymptotic expression for the Hankel functions:

$$\lim_{z\to\infty} H_n^{(1)}(z) = \sqrt{\frac{2}{\pi z}} e^{\hat{\imath}[z-(1+2n)\pi/4]}. \tag{56}$$

Incorporating Eq. (56) into the expression for the scattered wave in Eq. (22), the far-field scattered wave can be written as

$$\lim_{r\to\infty} \phi^{\mathrm{scr}} = \sqrt{\frac{2}{\pi kr}} e^{\hat{\imath}(kr-\pi/2)} \sum_{n=-\infty}^{\infty} B_n e^{\hat{\imath}n(\theta-\pi/2)}. \tag{57}$$

Note that the summation on the right-hand side of the above equation is a function of θ only, denoted as $f(\theta)$, and hence the name; that is,

$$f(\theta) = \sum_{n=-\infty}^{\infty} B_n e^{\hat{\imath}n(\theta-\pi/2)}. \tag{58}$$

Figures 11–13 compare the form factors for the three cases at three different frequencies: ka=2, 4, and 6 respectively. The three cases in each figure are the rigid cylinder and the cases of Fo=50 and 100. It is observed that with the cooling tube, the scattered wave pattern becomes prolonged in the forward direction, with most of the energy being directed into the forward direction. The pattern in the backward direction remains essentially unchanged. These figures provide a means of quantitative comparison to support the

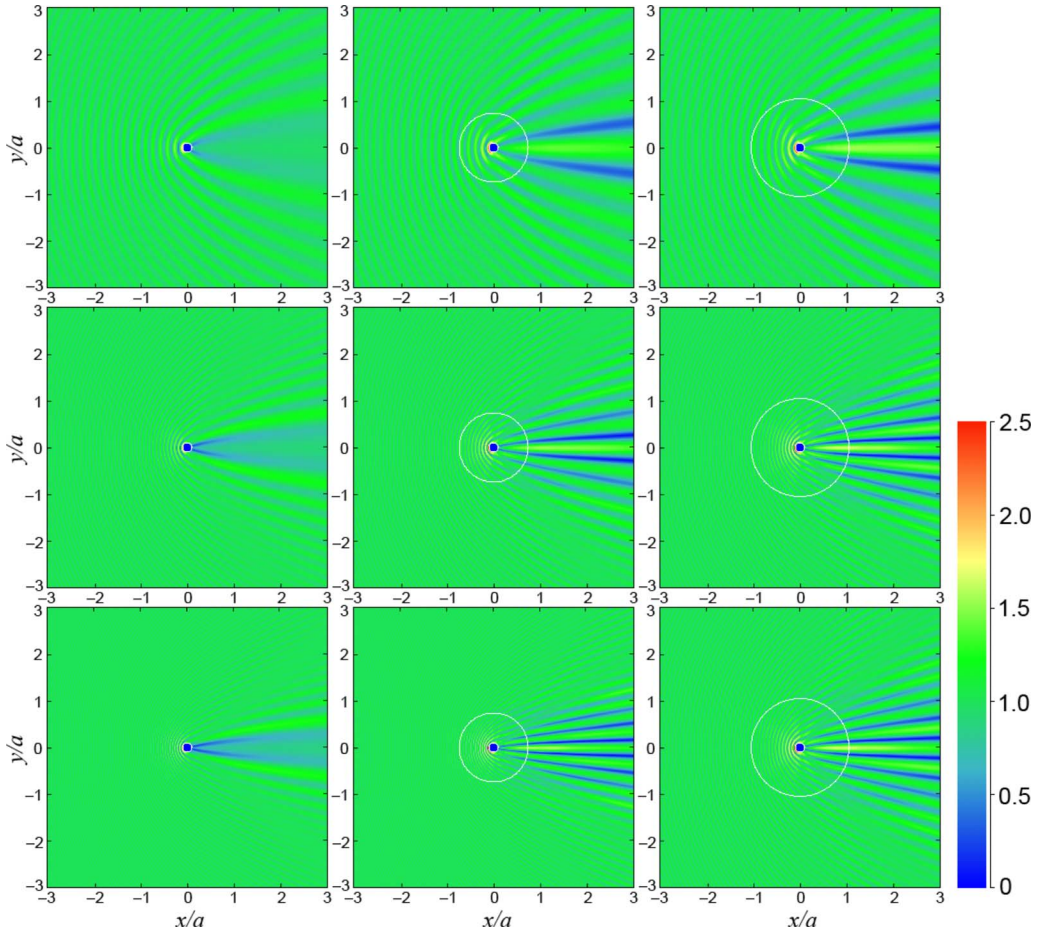


FIG. 10. (Color online) Pressure distributions in the vicinity of the cooling tube. Left column: a rigid cylinder, no cooling; center column: cooling tube with Fo=50; right column: cooling tube with Fo=100. First row: $ka=2$; second row: $ka=4$; bottom row: $ka=6$.

qualitative observations made in Fig. 10. A physical explanation is that the air surrounding the tube acts as a gradient-index lens in which the refractive index radially varies from a maximum value at the surface of the tube to a minimum value of 1 at the edge. This lens collects the sound and bends it around the tube, producing the characteristic temperature-dependent interference pattern.

D. Scattering cross section

The *total scattering cross section* is often used to describe the overall strength of a scatterer. It represents the energy flux (power) transmitted by the scattered wave across a closed surface in the far field.

The energy flux density averaged over one period of wave motion, called the *time-averaged energy flux density*, can be written as

$$\langle \dot{e} \rangle = \frac{\pi}{\omega} \Re\{\phi \bar{v}_n\}, \tag{59}$$

where v_n is the complex amplitude of the velocity component normal to the surface, ω is the frequency, and the over-

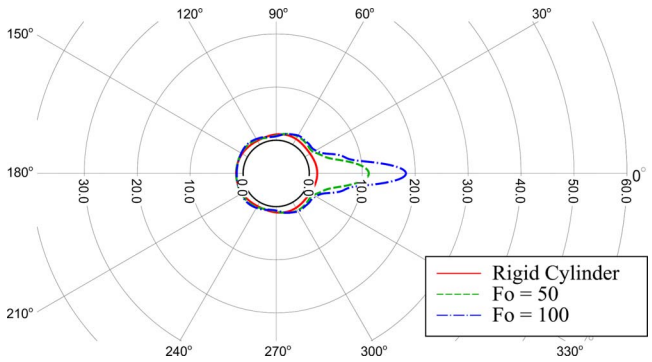


FIG. 11. (Color online) Form factor comparing the three cases at frequency $ka=2$.

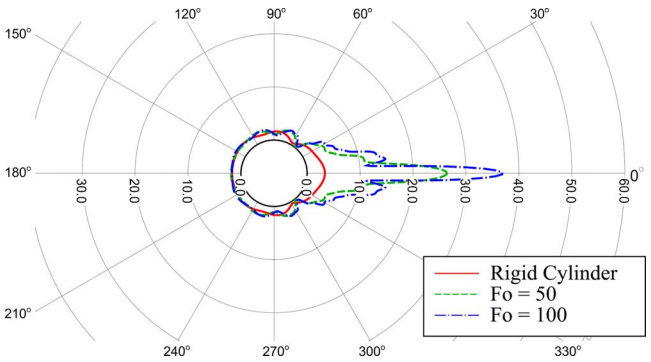


FIG. 12. (Color online) Form factor comparing the three cases at frequency $ka=4$.

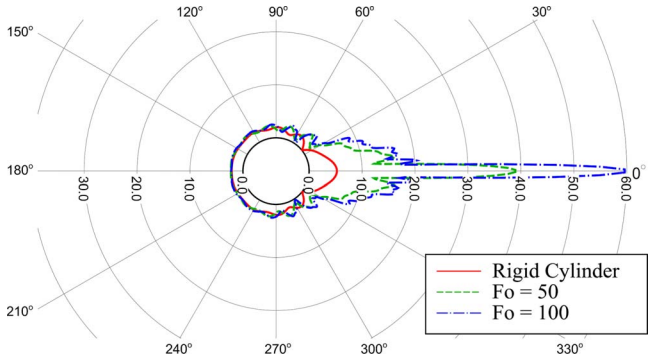


FIG. 13. (Color online) Form factor comparing the three cases at frequency $ka=6$.

bar denotes the complex conjugate. When the surface is a circular path,

$$v_n = v_r = -\frac{\hat{\imath}}{\omega\rho}\frac{\partial\phi}{\partial r}. \tag{60}$$

The energy flux density needs to be normalized by that of the incident wave. For a planar incident wave of the unit amplitude, its time-averaged energy flux density along its propagation direction is

$$\langle \dot{e}^{\text{inc}} \rangle = \frac{\pi}{\omega z}. \tag{61}$$

According to Eq. (59), the normalized time-averaged energy flux density of the scattered wave is

$$\langle \dot{\epsilon}^{\text{scr}} \rangle = z\Re\{\phi^{\text{scr}}\bar{v}_n^{\text{scr}}\}. \tag{62}$$

In the far field, the complex amplitude of radial velocity due to the scattered wave is

$$\lim_{r\to\infty} v_r^{\text{scr}} = \frac{f(\theta)}{z}\sqrt{\frac{2}{\pi kr}}e^{\hat{\imath}[kr-\pi/4]}. \tag{63}$$

Substituting Eqs. (57) and (63) into Eq. (62) gives

$$\lim_{r\to\infty}\langle \dot{\epsilon}^{\text{scr}} \rangle = \frac{2}{\pi kr}|f(\theta)|^2, \tag{64}$$

and integrating this quantity over the circular path of radius r gives the total scattering cross section as

$$\sigma = \frac{2}{\pi k}\int_0^{2\pi}|f(\theta)|^2 d\theta = \frac{4}{k}\{\boldsymbol{B}\}^*\{\boldsymbol{B}\}, \tag{65}$$

where, in the integration process, Eq. (58) has been used.

Figure 14 shows the spectra of the total scattering cross section for the three cases for frequencies up to $ka=10$. It can be seen that with cooling effect, the scattering cross section could be an order of magnitude larger than the rigid cylinder alone. The additional scattering strength appears to be mostly directed into the forward direction, as suggested by the form factor shown in Figs. 11–13. The difference between Fo=50 and the rigid cylinder (Fo=0) is larger than the difference between Fo=100 and Fo=50. The smoothness of the spectra corroborates with the observation that the phenomenon is consistently observable throughout the frequency range shown, except at very low frequencies.

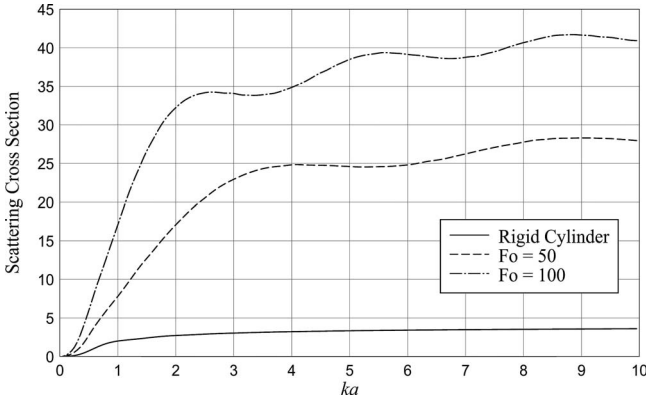


FIG. 14. Scattering cross section of three cases: rigid cylinder, Fo=50, and Fo=100.

Two other applications utilizing this procedure have been published elsewhere by the authors. One (Cai and Sánchez-Dehesa, 2007a) is similar to the present example and observes a focusing mechanism using multiple heated wires to form a grid. Another (Cai and Sánchez-Dehesa, 2007b) combines the present solution procedure with canonical solutions for acoustical scattering in mass-anisotropic metamaterials to perform an in-depth analysis of an acoustic cloaking (Cummer and Schurig, 2007).

VII. CONCLUSIONS

A recursive solution procedure has been proposed and validated for analyzing acoustical scattering by multilayer concentric circular cylindrical scatterers. The number of the layers can be arbitrary. The solution procedure is based on the concept of multiple scattering in a single scatterer. Computational characteristics, including potential pitfalls and their avoidance, are discussed. For a discrete multilayer system, the solution is analytically exact.

One important application of this solution procedure is for analyzing scatterers with radially stratified property profiles. Such scatterers can be approximated by a discrete multilayer system. In such cases, as the number of layers increases, the solution approaches the exact one.

A comprehensive numerical example uses the solution procedure to analyze the scattering of a cooling tube in air, which causes the surrounding air to become radially stratified. In this example, a part of the host medium is treated as a part of the scatterer. An interesting phenomenon is observed: when a planar wave encounters a cooling tube in space, instead of forming a shadow zone, a series of strongly collimated beam is formed behind the cooling tube.

ACKNOWLEDGMENTS

Most of this work is performed at the Universidad Politécnica de Valencia (UPV), Spain in the summer of 2007 when one of the authors (L.-W.C.) was a visiting professor there. The visit was sponsored by the National Science Foundation through an IREE supplemental funding of NSF Grant No. CMS-0510940 and by an *Ayuda para Estancia de Investigadores de Prestigio* grant from UPV. The program manager for the NSF base grant is Dr. C. V. Cooper, and the

program manager for the NSF IREE Program is Dr. A. Wing. L.-W.C. is grateful for the hospitality of his host, the co-author, and the entire Wave Phenomenon Group at UPV. J.S.-D. acknowledges the financial support from Spanish Ministerio de Educación y Ciencia (Grant No. TEC2007-67239) and Spanish Ministerio de Fomento (Grant No. C5/2006).

Barabás, M. (**1987**). "Scattering of a plane wave by a radially stratified tilted cylinder," J. Acoust. Soc. Am. **4**, 2240–2248.

Beranek, L. L. (**1986**). *Acoustics* (Acoustical Society of America, Melville, NY).

Bussey, H. W., and Richmond, J. H. (**1975**). "Scattering by a lossy dielectric circular cylindrical multilayer, numerical values," IEEE Trans. Antennas Propag. **AP30-23**, 723–725.

Cai, L.-W. (**2004a**). "Scattering of antiplane shear waves by layered circular elastic cylinder," J. Acoust. Soc. Am. **115**, 515–522.

Cai, L.-W. (**2004b**). "Multiple scattering in single scatterer," J. Acoust. Soc. Am. **115**, 986–995.

Cai, L.-W. (**2005**). "Scattering of elastic anti-plane shear wave by multilayered eccentric scatterers," Q. J. Mech. Appl. Math. **25**, 165–183.

Cai, L.-W., and Sánchez-Dehesa, J. (**2007a**). "Microlens array for focusing airborne ultrasond using heated wire grid," Appl. Phys. Lett. **91**, 181915.

Cai, L.-W., and Sánchez-Dehesa, J. (**2007b**). "Analysis of Cummer-Schurig acoustic cloaking," New J. Phys. **9**, 450.

Caorsi, S., and Pastorino, M. (**2004**). "Scattering my multilayer isorefractive elliptic cylinders," IEEE Trans. Antennas Propag. **52**, 189–196.

Carslaw, H. S., and Jaeger, J. C. (**1959**). *Conduction of Heat in Solids*, 2nd ed. (Oxford University Press, New York), pp. 256–261.

Chati, F., Leon, F., and Maze, G. (**2005**). "Acoustic scattering by a metallic tube with a concentric solid polymer cylinder coupled by a thin water layer. Influence of the thickness of the water layer on the two Scholte-Stoneley waves," J. Acoust. Soc. Am. **118**, 2820–2828.

Chew, W. C. (**1990**). *Waves and Fields in Inhomogeneous Media* (Van Nostrand Reinhold, New York), Chap. 3.

Chiu, C.-N., and Chen, C. H. (**1995**). "Plane-wave shielding properties of anisotropic laminated composite cylindrical shells," IEEE Trans. Electromagn. Compat. **39**, 109–113.

Cummer, S. A., and Schurig, D. (**2007**). "One path to acoustic cloaking," New J. Phys. **9**, 45.

Helaly, A., Soliman, E. A., and Megahed, A. A. (**1997**). "Electromagnetic waves scattering by nonuniform plasma cylinder," IEE Proc. Microwaves, Antennas Propag. **144**, 61–66.

Huang, W., Rokhlin, S. I., and Wang, Y. J. (**1995**). "Effect of fibre-matrix interphase on wave propagation along, and scattering from, multilayered fibres in composites. Transfer matrix method," Ultrasonics **33**, 365–375.

Ioannidou, M. P., Kapsalas, K. D., and Chrissoulidis, D. P. (**2004**). "Electromagnetic-wave scattering by an eccentrically stratified, dielectric cylinder with multiple, eccentrically stratified, cylindrical, dielectric inclusions," J. Electromagn. Waves Appl. **18**, 495–516.

Jiang, H., Han, X., and Li, R. (**2006**). "Improved algorithm for electromagnetic scattering of plane waves by a radially stratified tilted cylinder and its application," Opt. Commun. **266**, 13–18.

Kai, L., and D'Alessio, A. (**1995**). "Finely stratified cylinder model for radially inhomogeneous cylinders normally irradiated by electromagnetic plane waves," Appl. Opt. **34**, 5520–5530.

Kim, C. S., and Yeh, C. (**1991**). "Scattering of an obliquely incident wave by a multilayered elliptical lossy dielectric cylinder," Radio Sci. **26**, 1165–1176.

Kishk, A. A., Parrikar, R. P., and Elsherbeni, A. Z. (**1992**). "Electromagnetic scattering from an eccentric multilayered circular cylinder," IEEE Trans. Antennas Propag. **40**, 295–303.

Kluskens, M. S., and Newman, E. H. (**1991**). "Scattering by a multilayer chiral cylinder," IEEE Trans. Antennas Propag. **39**, 91–96.

Ragheb, H. A., and Hamid, M. (**1986**). "Scattering by two parallel multilayered dielectric cylinders," Int. J. Electron. **61**, 323–342.

Rao, T. C. K., and Hamid, M. A. K. (**1975**). "Scattering by a multi-layered dielectric-coated conducting cylinder," Int. J. Electron. **38**, 667–673.

Sastry, J. S., and Munjal, M. L. (**1998a**). "Response of a multi-layered infinite cylinder to a plane wave excitation by means of transfer matrices," J. Sound Vib. **209**, 99–121.

Sastry, J. S., and Munjal, M. L. (**1998b**). "Response of a multi-layered infinite cylinder to two-dimensional pressure excitation by means of transfer matrices," J. Sound Vib. **209**, 123–142.

Schurig, D., Mock, J. J., Justice, B. J., Cummer, S. A., Pendry, J. B., and Starr, A. F. (**2006**). "Metamaterial electromagnetic cloak at microwave frequencies," Science **314**, 977–980.

Skelton, E. A., and James, J. H. (**1993**). "Acoustics of an anisotropic layered cylinder," J. Sound Vib. **161**, 251–264.

Thomson, W. T. (**1950**). "Transmission of wave through a stratified solid medium," J. Appl. Phys. **21**, 89–93.

Torrent, D., and Sánchez-Dehesa, J. (**2007**). "Acoustic metamaterials for new two-dimensional sonic devices," New J. Phys. **9**, 323.

Waterman, P. C. (**1970**). "Symmetry, unitarity, and geometry in electromagnetic scattering," Phys. Rev. D **3**, 825–839.

Yeh, C., and Lindgren, G. (**1977**). "Computing the propagation characteristics of radially stratified fibers: An efficient method," Appl. Opt. **16**, 483–493.

Yin, W. Y., Li, L. W., and Leong, M. S. (**2000**). "Scattering from multiple bianisotropic cylinders and their modeling of cylindrical objects of arbitrary cross-section," Prog. Electromagn. Res. **27**, 159–184.

Yousif, H. A., and Elsherbeni, A. Z. (**1999**). "Electromagnetic scattering from an eccentric multilayered cylinder at oblique incidence," J. Electromagn. Waves Appl. **13**, 325–336.

Acoustic radiation torque on an irregularly shaped scatterer in an arbitrary sound field

Zongwei Fan, Deqing Mei,[a] Keji Yang, and Zichen Chen
Institute of Advanced Manufacturing Engineering, State Key Laboratory of Fluid Power Transmission and Control, Zhejiang University, Hangzhou 310027, People's Republic of China

(Received 5 February 2008; revised 28 July 2008; accepted 29 July 2008)

To eliminate the limitation of the conventional acoustic radiation torque theory, which is only applicable to a disklike scatterer in a plane sound field, a new theory is established to calculate the radiation torque on any irregularly shaped scatterer in any arbitrary sound field. First, with the aid of the conservation law of angular momentum, the acoustic radiation torque is expressed as the angular momentum flux through a spherical surface with the center at the scatterer's centroid. Second, the velocity potential of the scattered field is derived, taking into account the influences of the translational and rotational movements of the scatterer induced by the first order stress of the incident sound field. Finally, a general calculating formula of the acoustic radiation torque is achieved. For a disklike scatterer in a plane sound filed, results from the above formula are well identical with those conventional formulas. By studying the case of a semicircular cylinder scatterer in a standing-wave sound field, it is found that for an irregularly shaped scatterer its rotation velocity is normally nonzero and the radiation torque changes with the spatial attitude.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977733]

PACS number(s): 43.25.Qp [RMW] Pages: 2727–2732

I. INTRODUCTION

Acoustic radiation torque is a kind of nonlinear acoustic effect caused by the angular momentum transfer from a sound field to a scatterer. This time-averaged torque is the underlying principle of the noncontact rotational manipulation, which has been applied to control the orientation of fibers in a fiber-reinforced composite[1] and in a paper pulp.[2]

King studied the case of a circular disk in a plane progressive or stationary sound field.[3] Keller calculated the radiation torque caused by a plane progressive wave on a disk with a convex curvilinear rim,[4] and Maidanik derived a formula for an arbitrary shaped plane disk in a plane progressive field.[5]

Up to now, the theory of acoustic radiation torque is only applicable to a disklike scatterer in a plane sound field. However, the practical radiation field of an acoustic transducer cannot be simplified as a plane sound field, and other types of incident sound field besides a plane field may also be used, such as a synthesized field of two or more sound beams. Otherwise, scatterers may have various shapes; for example, the fibers in a fiber-reinforced composite and in a paper pulp should be treated as prolate spheroids. Therefore, the conventional theory has greatly restricted its applications, and a general theory applicable to any irregularly shaped scatterer in any arbitrary sound field needs to be investigated.

According to the conservation law of angular momentum, the acoustic radiation torque is equal to the time-averaged flux of the angular momentum through any closed surface enclosing the scatterer in the fluid. As a result, the key element to calculate the radiation torque is the scattered sound field, which is influenced by the oscillatory movements of the scatterer induced by the first order stress of the incident sound field. These movements include the translation of the mass center and the rotation of the scatterer's spatial attitude. In the traditional scattering theory, the spatial attitude rotation is not taken into account,[6] which is acceptable only when the local velocity of the incident sound field is always axisymmetric around one of the principal axes of the scatterer's induced-mass tensor. However, this axisymmetric assumption is fairly critical since the fluid element trajectory in a sound field is generally curvilinear and the scatterer's spatial attitude is also random. So the spatial attitude rotation should be taken into consideration for analyzing the scattered field.

In Sec. II, a theoretical model of the acoustic radiation torque is presented. In Sec. III, a generalized theory on Rayleigh scattering is obtained by considering the rotational movement. Section IV is about the derivation of the calculating formula for the acoustic radiation torque on an irregularly shaped scatterer in an arbitrary incident sound field. In Sec. V, three cases with different shaped scatterers in different incident sound fields are analyzed using the above theories. Section VI presents the conclusions.

All calculations in this paper are carried out in quadratic approximation with respect to the acoustic Mach number, which is defined as the ratio of the fluid velocity to the sound speed. Moreover, the time dependence of the first order fields and their derivatives is assumed to be periodic.

II. THEORETICAL MODEL OF ACOUSTIC RADIATION TORQUE

Acoustic radiation torque is induced by the angular momentum transfer from sound fields to scattering objects. The

[a]Author to whom correspondence should be addressed. Electronic mail: medqmei@zju.edu.cn

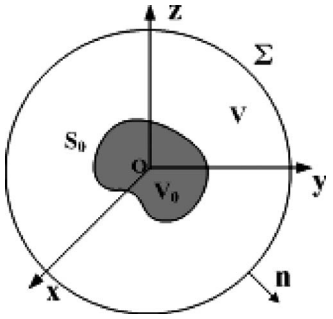


FIG. 1. Coordinate system for the calculation of the acoustic radiation torque on an arbitrary shape scatterer. The origin point O is at the centroid of the scatterer.

conservation equation for angular momentum can be derived from that for linear momentum, which is

$$\frac{\partial \rho u_k}{\partial t} + \frac{\partial}{\partial x_l}\Pi_{kl} = 0, \tag{1}$$

where ρ and u_k are the fluid density and velocity, respectively, and Π_{kl} is the momentum flux density tensor.[6] Let r_j be a position vector, so that Eq. (1) transforms into

$$\varepsilon_{ijk} r_j \frac{\partial \rho u_k}{\partial t} + \varepsilon_{ijk} r_j \frac{\partial}{\partial x_l}\Pi_{kl} = 0, \tag{2}$$

where ε_{ijk} is the permutation tensor. Equation (2) can be rewritten as

$$\frac{\partial}{\partial t}(\varepsilon_{ijk} r_j \rho u_k) + \frac{\partial}{\partial x_l}(\varepsilon_{ijk} r_j \Pi_{kl}) = 0, \tag{3}$$

where $\varepsilon_{ijk} r_j \rho u_k$ is the angular momentum density and $\varepsilon_{ijk} r_j \Pi_{kl}$ is the angular momentum flux density tensor. Equation (3) is exactly the conservation equation for angular momentum.

According to the conservation law of angular momentum, the acoustic radiation torque exerted on a scatterer by an incident sound field is equal to the time-averaged flux of the angular momentum through any closed surface in the fluid that encloses the scatterer,

$$T_i = \left\langle -\oiint_\Sigma \varepsilon_{ijk} r_j \Pi_{kl} n_l dS \right\rangle, \tag{4}$$

where pointed brackets denote the time-averaged operation, Σ is the integral surface, and n_j is the outward normal of Σ. Let Σ be a spherical surface, with its center at the centroid of the scatterer, as shown in Fig. 1, in which S_0 is the exterior surface of the scatterer, V_0 is the volume occupied by the scatterer, and V is the volume bounded by Σ and S_0.

If the scatterer's size is much greater than the penetration depth of the viscous wave, the influence of the acoustic streaming and the viscosity of the medium can be neglected. In this case, the momentum flux density tensor is

$$\Pi_{kl} = p\delta_{kl} + \rho_0 U_k U_l, \tag{5}$$

where p is the fluid pressure, U_k is the fluid velocity, ρ_0 is the equilibrium density of the fluid medium, and δ_{kl} is the Kronecker delta.

Let the potential functions of the incident and scattered fields be denoted by φ_{in} and φ_{sc}, respectively. The potential function of the total sound field is $\varphi = \varphi_{in} + \varphi_{sc}$, and the fluid velocity is $U = \nabla\varphi$. Correcting to terms of the order of U^2/c_0^2, the pressure variation in the fluid medium is

$$p = -\rho_0 \frac{\partial \varphi}{\partial t} - \rho_0 \frac{U^2}{2} + \frac{\rho_0}{2c_0^2}\left(\frac{\partial \varphi}{\partial t}\right)^2, \tag{6}$$

where c_0 is the velocity of sound in the fluid medium.[7]

Because of the periodicity of $\partial\varphi/\partial t$, its time-averaged value equals zero. Substituting Eqs. (5) and (6), Eq. (4) can be rewritten as

$$T_i = \left\langle \oiint_\Sigma \varepsilon_{ijk} r_j \left\{ \left[\rho_0 \frac{U^2}{2} - \frac{\rho_0}{2c_0^2}\left(\frac{\partial \varphi}{\partial t}\right)^2 \right] \delta_{kl} - \rho_0 U_k U_l \right\} n_l dS \right\rangle. \tag{7}$$

In Eq. (7), the incident field is already known, so the scattered field is the key to calculate the acoustic radiation torque.

III. SCATTERED FIELD OF AN IRREGULARLY SHAPED RAYLEIGH SCATTERER

A. Movements of the scatterer induced by the incident sound field

Under the action of the first order stress of the incident sound field, the scatterer will execute an oscillatory motion, including the mass center translation and the spatial attitude rotation. The total action includes two parts: The first part is the action on the scatterer when it is wholly carried by the fluid and equals the action on the fluid in the same volume as that occupied by the scatterer. The second part is the reaction due to the scatterer's motion related to the fluid when the scatterer is not wholly carried along with the fluid. By applying the expression of the reactive force and couple,[8] the equations to calculate the movements of the scatterer are

$$\rho_s V_0 \frac{du_i}{dt} = \rho_0 V_0 \frac{dv_i}{dt} - \frac{dP_i}{dt} + \varepsilon_{ijk} P_j \omega_k,$$

$$I_i \frac{d\omega_i}{dt} = -\frac{dQ_i}{dt} + \varepsilon_{ijk} P_j (u_k - v_k) + \varepsilon_{ijk} Q_j \omega_k, \tag{8}$$

where u_i and ω_i are the translating and rotating velocities of the scatterer, respectively, v_i is the local velocity of the incident sound field, ρ_s is the material density of the scatterer, V_0 is the volume of the scatterer, I_i is the ith coordinate axis inertia moment of the scatterer, and P_i and Q_i are the fluid momentum and angular momentum induced by the scatterer's movements relative to the fluid, respectively. The definitions of the latter are

$$P_i = \sum_{j=1}^{3} [\lambda_{ij}(u_j - v_j) + \lambda_{i,j+3}\omega_j], \tag{9a}$$

$$Q_i = \sum_{j=1}^{3} [\lambda_{i+3,j}(u_j - v_j) + \lambda_{i+3,j+3}\omega_j], \tag{9b}$$

where λ is the element of the scatterer's induced-mass tensor and $i,j=1,2,3$.

After linearization, Eq. (8) can be rewritten as

$$\rho_s V_0 \frac{du_i}{dt} = \rho_0 V_0 \frac{dv_i}{dt} - \frac{dP_i}{dt},$$

$$I_i \frac{d\omega_i}{dt} = -\frac{dQ_i}{dt}. \tag{10}$$

Since u_i and ω_i will vanish when v_i vanishes, by integrating both sides of Eq. (10) with respect to time, it will change into

$$\rho_s V_0 u_i = \rho_0 V_0 v_i - P_i,$$

$$I_i \omega_i = -Q_i. \tag{11}$$

Based on Eq. (11), u_i and ω_i can be calculated eventually.

B. Potential function of the scattered field

Since the size of the scatterer is much smaller than the wavelength of the sound wave, the potential function of the scattered field can be expressed as[6]

$$\varphi_{sc} = -\frac{V_0}{4\pi r}\frac{1}{\rho_0}\dot{\rho}_{in}(t-r/c_0)\left(1-\frac{c_0^2\rho_0}{c_s^2\rho_s}\right) + \frac{\partial}{\partial x_i}\left[\frac{A_i(t-r/c_0)}{r}\right], \tag{12}$$

where c_s is the sound speed of the scatterer's material, $\dot{\rho}_{in}(t-r/c_0)$ is the differential of the fluid density function of the incident sound field with respect to time, r denotes the radial distance from the original point O, the vector function A_i is

$$A_i = \frac{P_i + \rho_0 V_0 (u_i - v_i)}{4\pi\rho_0}. \tag{13}$$

The verification of Eq. (13) is as follows. Let the fluid velocity of the scattered field be denoted by w_i, then

$$\begin{aligned}\iiint_V w_i w_i d\tau &= \iiint_V [w_i + (u_i - v_i) + \varepsilon_{ijk}\omega_j r_k] \times [w_i - (u_i - v_i) - \varepsilon_{ijk}\omega_j r_k] d\tau \\ &+ \iiint_V [(u_i - v_i) + \varepsilon_{ijk}\omega_j r_k][(u_i - v_i) + \varepsilon_{ijk}\omega_j r_k] d\tau = \iiint_V \frac{\partial}{\partial x_i}\{[\varphi_{sc} + (u_j - v_j) r_j] \\ &\times [w_i - (u_i - v_i) - \varepsilon_{ijk}\omega_j r_k]\} d\tau + \iiint_V (u_i - v_i)(u_i - v_i) d\tau \\ &+ \iiint_V \varepsilon_{ijk}\omega_j r_k [w_i + (u_i - v_i)] d\tau. \end{aligned} \tag{14}$$

Using the Gauss theorem and the nonpenetrating boundary condition on the exterior surface of the scatterer, i.e., $[w_i - (u_i - v_i) - \varepsilon_{ijk}\omega_j r_k]n_i = 0$ on S_0, Eq. (14) can be transformed into

$$\begin{aligned}\iiint_V w_i w_i d\tau &= \iint_\Sigma [\varphi_{sc} + (u_j - v_j) r_j] \times [w_i - (u_i - v_i)] n_i dS + \iiint_V (u_i - v_i) \\ &\times (u_i - v_i) d\tau + \iiint_V \varepsilon_{ijk}\omega_j r_k [w_i + (u_i - v_i)] d\tau \\ &= \int \left[-\frac{A_i n_i}{R^2} + (u_i - v_i) n_i R\right]\left[\frac{3A_i n_i - A_i n_i}{R^3} - (u_i - v_i) n_i\right] R^2 do + \left(\frac{4}{3}\pi R^3 - V_0\right)(u_i - v_i) \\ &\times (u_i - v_i) + \iiint_V \varepsilon_{ijk}\omega_j r_k [w_i + (u_i - v_i)] d\tau, \end{aligned} \tag{15}$$

where do is the differential of the solid angle. Using the vector identity $\iiint_V \varepsilon_{ijk}\omega_j r_k w_i d\tau = \iiint_V \omega_i \varepsilon_{ijk} r_j w_k d\tau = \omega_i \iiint_V \varepsilon_{ijk} r_j w_k d\tau$, Eq. (15) can be rewritten as

$$\begin{aligned}\iiint_V w_i w_i d\tau &= 4\pi A_i (u_i - v_i) - V_0 (u_i - v_i)(u_i - v_i) \\ &+ \omega_i \iiint_V \varepsilon_{ijk} r_j w_k d\tau \\ &+ \iiint_V (u_i - v_i)\varepsilon_{ijk}\omega_j r_k d\tau. \end{aligned} \tag{16}$$

According to the definition of the angular momentum Q_i, the third term on the right of Eq. (16) is

$$\omega_i \iiint_V \varepsilon_{ijk} r_j w_k d\tau = \frac{Q_i \omega_i}{\rho_0}. \tag{17}$$

Since the original point O is the centroid of the scatterer and also the sphere bounded by Σ, it can be derived that point O is exactly the centroid of the volume V bounded by Σ and S_0. Thus, according to the theory of classical mechanics,[9] the fourth term on the right of Eq. (16) is

$$\iiint_V (u_i - v_i)\varepsilon_{ijk}\omega_j r_k d\tau = 0. \tag{18}$$

Based on the analysis above, Eq. (14) is equal to

$$\iiint_V w_i w_i d\tau = 4\pi A_i (u_i - v_i) - V_0 (u_i - v_i)(u_i - v_i) + \frac{Q_i \omega_i}{\rho_0}. \tag{19}$$

On the other hand, from the definition of the kinematical energy of fluid induced by a moving object, the following equation can be obtained:

$$\iiint_V w_i w_i d\tau = \frac{P_i (u_i - v_i) + Q_i \omega_i}{\rho_0}. \tag{20}$$

Combining Eqs. (19) and (20), the expression for the vector function A_i, i.e., Eq. (13), is derived.

IV. ACOUSTIC RADIATION TORQUE ON AN IRREGULARLY SHAPED SCATTERER IN AN ARBITRARY SOUND FIELD

Using the Gauss theorem, the expression of the acoustic radiation torque [Eq. (4)] can be transformed into a volume integral,

$$\begin{aligned} T_i &= \left\langle -\oiint_{\Sigma} n_l (r_{i+1}\Pi_{i+2,l} - r_{i+2}\Pi_{i+1,l} dS) \right\rangle \\ &= \left\langle -\iiint_{V+V_0} \frac{\partial}{\partial x_l}(r_{i+1}\Pi_{i+2,l} - r_{i+2}\Pi_{i+1,l}) d\tau \right\rangle \\ &= \left\langle -\iiint_{V+V_0} \left(r_{i+1}\frac{\partial \Pi_{i+2,l}}{\partial x_l} - r_{i+2}\frac{\partial \Pi_{i+1,l}}{\partial x_l} \right) d\tau \right\rangle. \end{aligned} \tag{21}$$

Equation (21) can be divided into two parts: One is from the interference between the incident and scattered fields, and the other is from the self-interference of the scattered field.

First, the contribution to the radiation torque from the interference of the incident and scattered fields is calculated. The part of the momentum flux density caused by the interference terms is

$$\Pi_{kl} = \left(-\rho_0 v_m w_m + \frac{c_0^2}{\rho_0}\rho_{\text{in}}\rho_{\text{sc}} \right)\delta_{kl} + \rho_0(v_k w_l + v_l w_k). \tag{22}$$

Substituting Eq. (22) into Eq. (21), the following equation can be gained:

$$\begin{aligned} \frac{\partial \Pi_{il}}{\partial x_l} &= \rho_0 v_i \left(\nabla^2 \varphi_{\text{sc}} - \frac{1}{c_0^2}\frac{\partial^2 \varphi_{\text{sc}}}{\partial t^2} \right) + \rho_0 w_i \left(\nabla^2 \varphi_{\text{in}} - \frac{1}{c_0^2}\frac{\partial^2 \varphi_{\text{in}}}{\partial t^2} \right) \\ &= \rho_0 v_i \left(\nabla^2 \varphi_{\text{sc}} - \frac{1}{c_0^2}\frac{\partial^2 \varphi_{\text{sc}}}{\partial t^2} \right). \end{aligned} \tag{23}$$

In Eq. (23), under the operation of the d'Alembert operator, the first term on the right of Eq. (12) transfers to

$$\begin{aligned} &\left(\nabla^2 - \frac{1}{c_0^2}\frac{\partial^2}{\partial t^2} \right)\left[-\frac{V_0}{4\pi r}\frac{1}{\rho_0}\dot{\rho}_{\text{in}}(t - r/c_0)\left(1 - \frac{c_0^2\rho_0}{c_s^2\rho_s} \right) \right] \\ &= -\frac{V_0}{\rho_0}\left(1 - \frac{c_0^2\rho_0}{c_s^2\rho_s} \right)\dot{\rho}_{\text{in}}(t)(\nabla^2 + k^2)\left(\frac{e^{ikr}}{4\pi r} \right) \\ &= \frac{V_0}{\rho_0}\left(1 - \frac{c_0^2\rho_0}{c_s^2\rho_s} \right)\dot{\rho}_{\text{in}}(t)\,\delta(r), \end{aligned} \tag{24}$$

where $\delta(r)$ is the Dirac function.[10] Similarly, the second term of φ_{sc} transfers to

$$\begin{aligned} &\left(\nabla^2 - \frac{1}{c_0^2}\frac{\partial^2}{\partial t^2} \right)\left\{ \frac{\partial}{\partial x_i}\left[\frac{A_i(t - r/c_0)}{r} \right] \right\} \\ &= -4\pi \frac{\partial}{\partial x_i}[A_i(t)\,\delta(r)]. \end{aligned} \tag{25}$$

Substituting Eqs. (23)–(25) into Eq. (21), the first part of the acoustic radiation torque

$$\begin{aligned} T_i &= \left\langle -\rho_0 \iiint_{V+V_0} (r_{i+1}v_{i+2} - r_{i+2}v_{i+1}) \right. \\ &\quad \left. \times \left(\nabla^2 \varphi_{\text{sc}} - \frac{1}{c_0^2}\frac{\partial^2 \varphi_{\text{sc}}}{\partial t^2} \right) d\tau \right\rangle \\ &= \left\langle -\rho_0 \iiint_{V+V_0} (r_{i+1}v_{i+2} - r_{i+2}v_{i+1}) \right. \\ &\quad \left. \times \left\{ \frac{V_0}{\rho_0}\left(1 - \frac{c_0^2\rho_0}{c_s^2\rho_s} \right)\dot{\rho}_{\text{in}}(t)\,\delta(r) - 4\pi \frac{\partial}{\partial x_j}[A_j(t)\,\delta(r)] \right\} d\tau \right\rangle \\ &= \left\langle 4\pi\rho_0 A_j(t) \iiint_{V+V_0} (r_{i+1}v_{i+2} - r_{i+2}v_{i+1}) \frac{\partial\,\delta(r)}{\partial x_j} d\tau \right\rangle \\ &= \left\langle -4\pi\rho_0 A_j(t) \iiint_{V+V_0} \delta(r) \frac{\partial (r_{i+1}v_{i+2} - r_{i+2}v_{i+1})}{\partial x_j} d\tau \right\rangle \\ &= -4\pi\rho_0 \langle A_{i+1}(t)v_{i+2} - A_{i+2}(t)v_{i+1} \rangle. \end{aligned} \tag{26}$$

Second, the second contribution to Eq. (21) from the self-interference of the scattered field is calculated as follows:

$$\begin{aligned} T_i &= \left\langle -\rho_0 \oiint_{\Sigma} n_l w_l \varepsilon_{ijk} r_j w_k dS \right\rangle \\ &= \left\langle -\rho_0 \int_0^{2\pi}\int_0^{\pi} \left\{ \frac{V_0 \ddot{\rho}_{\text{in}}(t - r/c_0)}{4\pi\rho_0 c_0 r} + \frac{V_0 \dot{\rho}_{\text{in}}(t - r/c_0)}{4\pi\rho_0 r^2} \right. \right. \\ &\quad + \left[\frac{\ddot{A}_l(t - r/c_0)}{c_0^2 r} + \frac{\dot{A}_l(t - r/c_0)}{c_0 r^2} \right. \\ &\quad \left. \left. + \frac{2A_l(t - r/c_0)}{r^3} \right] n_l \right\} \varepsilon_{ijk} r n_j \left[\left(-\frac{\dot{A}_l}{c_0 r^2}\frac{\partial n_l}{\partial \psi} - \frac{A_l}{r^3}\frac{\partial n_l}{\partial \psi} \right) \right. \\ &\quad \times (e_\psi)_k + \left(-\frac{\dot{A}_l}{c_0 r^2 \sin\psi}\frac{\partial n_l}{\partial \varphi} - \frac{A_l}{r^3 \sin\psi}\frac{\partial n_l}{\partial \varphi} \right) \\ &\quad \left. \left. \times (e_\varphi)_k \right] r^2 \sin\psi d\psi d\varphi \right\rangle \\ &= \left\langle -\rho_0 \int_0^{2\pi}\int_0^{\pi} \left\{ \frac{V_0 \ddot{\rho}_{\text{in}}(t - r/c_0)}{4\pi\rho_0 c_0 r} + \frac{V_0 \dot{\rho}_{\text{in}}(t - r/c_0)}{4\pi\rho_0 r^2} \right. \right. \\ &\quad + \left[\frac{\ddot{A}_l(t - r/c_0)}{c_0^2 r} + \frac{\dot{A}_l(t - r/c_0)}{c_0 r^2} + \frac{2A_l(t - r/c_0)}{r^3} \right] n_l \Bigg\} \\ &\quad \times \left[\left(-\frac{\dot{A}_l(e_\psi)_l}{c_0 r^2} - \frac{A_l(e_\psi)_l}{r^3} \right)(e_\varphi)_i + \left(\frac{\dot{A}_l(e_\varphi)_l}{c_0 r^2} + \frac{A_l(e_\varphi)_l}{r^3} \right) \right. \\ &\quad \left. \left. \times (e_\psi)_i \right] r^3 \sin\psi d\psi d\varphi \right\rangle \\ &= \left\langle -\rho_0 \int_0^{2\pi}\int_0^{\pi} \left[\frac{\ddot{A}_l(t - r/c_0)}{c_0^2 r} + \frac{\dot{A}_l(t - r/c_0)}{c_0 r^2} \right. \right. \\ &\quad \left. \left. + \frac{2A_l(t - r/c_0)}{r^3} \right] n_l \left[\left(-\frac{\dot{A}_l(e_\psi)_l}{c_0 r^2} - \frac{A_l(e_\psi)_l}{r^3} \right)(e_\varphi)_i \right. \right. \end{aligned}$$

$$+\left(\frac{\dot{A}_l(e_\varphi)_l}{c_0 r^2}+\frac{A_l(e_\varphi)_l}{r^3}\right)(e_\psi)_i\Bigg]r^3\sin\psi d\psi d\varphi\Bigg\rangle$$

$$=\Bigg\langle-\frac{\rho_0}{c_0^3}\int_0^{2\pi}\int_0^{\pi}\ddot{A}_l(t-r/c_0)n_l[-\dot{A}_l(e_\psi)_l(e_\varphi)_i$$

$$+\dot{A}_l(e_\varphi)_l(e_\psi)_i]\sin\psi d\psi d\varphi\Bigg\rangle=-\frac{8\pi\rho_0}{3c_0^3}\langle\dot{A}_{i+1}\ddot{A}_{i+2}\rangle, \tag{27}$$

where the dot on top of a character denotes the first derivative with respect to time, two dots means the second derivative, r is the radius of the sphere bounded by the surface Σ, ψ and φ are colatitude and azimuth variables in a spherical coordinate system with the origin at point O, respectively, and e_ψ and e_φ are unit vectors along the colatitude and azimuth directions, respectively.

From Eqs. (26) and (27), the following conclusions can be achieved as follows.

(a) If the trajectory of the local fluid velocity is rectilinear, it can be proven that the second part of the radiation torque equals zero strictly.
(b) If the trajectory is not linear, the second part of the radiation torque is normally nonzero. However, it can also be proven that its magnitude order is the same as that of $\rho_0 V_0 v^2(ka)^3$, while the first has the same magnitude order with $\rho_0 V_0 v^2$. Therefore, the contribution of the self-interference of the scattered field to the acoustic radiation torque is negligible.
(c) Besides the acoustic radiation torque, there is another type of nonlinear steady torque induced by the near-boundary acoustic streaming in the vicinity of a scatterer.[11] A synthesized field of two standing waves, with the same frequency vibrating out of phase, has been used to exert a steady torque on a spherical or circular cylindrical scatterer. For irregularly shaped scatterers, acoustic radiation torque occurs and will compete with the torque induced by the streaming. Referring to the expression for the steady torque induced by the near-boundary streaming in Ref. 12, it can be seen that the magnitude order of the torque of the streaming is the same as that of $\rho_0 V_0 v^2(\zeta_v/a)$, in which ζ_v is the penetration depth of the viscous wave. Hence, by comparing with the acoustic radiation torque, the torque from the acoustical streaming can also be neglected if ζ_v is much smaller than the size of the scatterer.

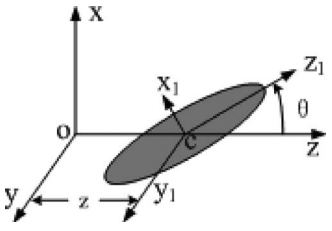


FIG. 2. Coordinate system for the calculation of the acoustic radiation torque on a spheroidal scatterer in a standing-wave sound field.

V. APPLICATION

A. Arbitrarily shaped disk in a plane progressive wave

In 1958, Maidanik derived a formula for the acoustic radiation torque on a plane disk of arbitrary shape in a plane progressive wave.[5] The induced-mass tensor of a plane disk is

$$[\lambda_{ij}]=\begin{pmatrix}0&0&0\\0&0&0\\0&0&\lambda_{zz}\end{pmatrix}. \tag{28}$$

Based on Eq. (28) and the theory established in Secs. III and IV, the acoustic radiation torque exerted by a plane progressive wave is

$$T_x=T_z=0,$$

$$T_y=\tfrac{1}{4}k^2\lambda_{zz}\sin 2u, \tag{29}$$

where k is the wave number and u is the angle between the incident direction and the normal of the disk. Equation (29) is identical with Eqs. (41) and (42) in Ref. 5.

B. Spheroid in a plane standing-wave field

Suppose that a spheroidal scatterer is placed in a plane standing-wave field, as shown in Fig. 2. Let the center of the spheroid on the z axis and its coordinate be $(0,0,z)$. The angle between the x_1 axis and the z axis is denoted as θ. If the axis z_1 is the symmetrical axis of this spheroidal scatterer, the induced-mass in the x_1 and y_1 directions are both $k_1\rho_0 V_0$, and the induced-mass in the z_1 direction is $k_3\rho_0 V_0$. Applying the theory derived above, the acoustic radiation torque on a spheroidal scatterer in a plane standing-wave field is equal to

$$T_x=T_z=0,$$

$$T_y=\frac{1}{4}\rho_0 V_0 v_0^2\sin^2 kz\sin 2\theta\Bigg[\left(\frac{1+k_3}{\rho_s+k_3\rho_0}-\frac{1+k_1}{\rho_s+k_1\rho_0}\right)\times(\rho_0-\rho_s)\Bigg]. \tag{30}$$

In the processes of preparing the fiber-reinforced composite and paper making, the fiber is assumed to be a prolate ellipsoidal scatterer; therefore $k_1>k_3$. If the density $\rho_0<\rho_s$, the coefficient in the square brackets of Eq. (30) will always be positive. In this case, the scatterer has a stable spatial attitude $\theta=\pi/2$. This can explain why the fibers in a standing-wave sound field have a balancing orientation parallel to the pressure node planes.[1,2]

C. Semicircular cylinder in a plane standing-wave field

Let a semicircular cylinder with radius R be placed in a plane standing-wave field. The axis of the cylinder is parallel

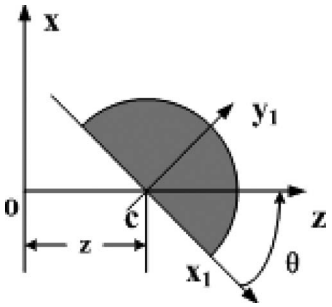


FIG. 3. Coordinate system for the calculation of the acoustic radiation torque on a semicircular cylinder scatterer in a standing-wave sound field.

to the pressure node planes, the circular center's coordinate is $(0,0,z)$, and the angle between the x_1 and z axes is θ, as shown in Fig. 3.

Using the boundary element method, the induced-mass tensor can be calculated as

$$[\lambda_{ij}]_{3\times 3} = \rho_0 R^2 \begin{pmatrix} 1.57 & 0 & -0.67\cos\theta \\ 0 & 1.57 & 0.67\sin\theta \\ -0.67\cos\theta & 0.67\sin\theta & 0.49 \end{pmatrix}. \tag{31}$$

Substituting Eq. (31) into Eq. (11), we obtain

$$\omega_y = \frac{1.57^{*}0.67\cos\theta(\zeta-1)}{0.67^2 - 1.57(0.49+0.79\zeta)(1+\zeta)} v_z, \tag{32}$$

where $\zeta=\rho_s/\rho_0$. From Eq. (32), it is found that the angular velocity ω_y is zero only when $\zeta=1$ or $\theta=\pi/2$, so the traditional scattering theory is not rigorous to neglect the spatial attitude rotation in most cases. Based on Eq. (32) and the theories above, the following equation can be obtained:

$$T_x = T_z = 0,$$

$$T_y = \frac{1}{4}\rho_0 R^2 v_0^2 \sin^2 kz \sin 2\theta \frac{1.57 * 0.67^2 * (\zeta-1)^2/(\zeta+1)}{0.67^2 - 1.57(0.49+0.79\zeta)(1+\zeta)}. \tag{33}$$

From Eq. (33), it can be seen that the semicircular cylinder has a stable balancing attitude $\theta=0$ in a standing-wave field because the coefficient of $\sin 2\theta$ in Eq. (33) is always negative.

VI. CONCLUSIONS

A general theory of the acoustic radiation torque on an irregular Rayleigh scatterer in an arbitrary sound field has been established. Based on the conservation law of angular momentum, the acoustic radiation torque is expressed as the time-averaged flux of the angular momentum through a spherical surface with the center at the scatterer's centroid.

The traditional theory of Rayleigh scattering has been generalized by considering the translational and rotational movements of the scatterer induced by the first order stress of the incident sound field.

The acoustic radiation torque is divided into two parts: one is from the interference between the incident and scattered fields, and the other is from the self-interference of the scattered field. By substituting the scattered field, these two parts of the radiation torque are calculated. By comparing with the acoustic radiation torque, it is proven that the steady torque induced by the near-boundary acoustical streaming is negligible if the penetration depth of the viscous wave is much smaller than the size of the scatterer.

For a disklike scatterer in a plane sound filed, results calculated by the formula above are identical with conventional formulas. By studying the case of a semicircular cylinder scatterer in a standing-wave sound field, it is found that the rotation velocity is normally nonzero and the radiation torque changes with its spatial attitude for an irregularly shaped scatterer.

The research results obtained in this work provide the theoretical fundament to the correlative applications, such as the noncontact rotation and the attitude control of microobjects. Moreover, the knowledge of the balancing attitude of the scatterer under the action of the acoustic radiation torque is critical to calculate the acoustic radiation force on an irregularly shaped scatterer. By combining the acoustic radiation torque and the radiation force, the noncontact 6DOF micromanipulation technology will be anticipated.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support from Hi-Tech Research and Development Program of China (Grant No. 2006AA04Z329), National Natural Science Foundation of China (Grant No. 50775203). The first author (Z.W.F.) acknowledges Professor Yaodong Chen and Professor Bingfeng Ju from Zhejiang University for valuable discussions.

[1] S. Yamahira, S. Hatanaka, M. Kuwabara, and S. Asai, "Orientation of fibers in liquid by ultrasonic standing waves," Jpn. J. Appl. Phys., Part 1 **39**, 3683–3687 (2000).
[2] P. Brodeur, J. L. Dion, and P. L. H, "Reorientation and layering of fibers in stationary ultrasonic waves," Proceedings of the IEEE Ultrasonics Symposium, Denver, CO (14–16 October 1987).
[3] L. V. King, "On the theory of the inertia and diffraction corrections for the Rayleigh disc," Proc. R. Soc. London, Ser. A **153**, 17–40 (1935).
[4] J. B. Keller, "Acoustic torques and forces on disks," J. Acoust. Soc. Am. **29**, 1085–1090 (1957).
[5] G. Maidanik, "Torques due to acoustical radiation pressure," J. Acoust. Soc. Am. **30**, 620–623 (1958).
[6] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1987).
[7] L. V. King, "On the acoustic radiation pressure on spheres," Proc. R. Soc. London, Ser. A **147**, 212–240 (1935).
[8] L. M. Milne-Thomson, *Theoretical Hydrodynamics* (Macmillan, London, 1968).
[9] H. Goldstein, C. Poole, and J. Safko, *Classical Mechanics* (Addison-Wesley, New York, 2002).
[10] P. H. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[11] F. H. Busse and T. G. Wang, "Torque generated by orthogonal acoustic waves—theory," J. Acoust. Soc. Am. **69**, 1634–1687 (1981).
[12] C. P. Lee and T. G. Wang, "Near-boundary streaming around a small sphere due to two orthogonal standing waves," J. Acoust. Soc. Am. **85**, 1081–1088 (1989).

A resonance shift prediction based on the Boltzmann–Ehrenfest principle for cylindrical cavities with a rigid sphere

Arturo O. Santillan[a)]
Centro de Ciencias Aplicadas y Desarrollo Tecnológico, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Cd. Universitaria, A. P. 70-186, 02510 México, Distrito Federal, México

Vicente Cutanda-Henríquez
Institute of Sensors, Signals and Electrotechnics, Faculty of Engineering, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

(Received 17 March 2008; revised 7 August 2008; accepted 19 August 2008)

An investigation on the resonance frequency shift for a plane-wave mode in a cylindrical cavity produced by a rigid sphere is reported in this paper. This change of the resonance frequency has been previously considered as a cause of oscillational instabilities in single-mode acoustic levitation devices. It is shown that the use of the Boltzmann–Ehrenfest principle of adiabatic invariance allows the derivation of an expression for the resonance frequency shift in a simpler and more direct way than a method based on a Green's function reported in literature. The position of the sphere can be any point along the axis of the cavity. Obtained predictions of the resonance frequency shift with the deduced equation agree quite well with numerical simulations based on the boundary element method. The results are also confirmed by experiments. The equation derived from the Boltzmann–Ehrenfest principle appears to be more general, and for large spheres, it gives a better approximation than the equation previously reported. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2982422]

PACS number(s): 43.25.Uv, 43.20.Ks, 43.20.Fn [OAS] Pages: 2733–2741

I. INTRODUCTION

Forces generated by acoustic waves can compensate for the weight of small samples, which can therefore be suspended in a fluid. This phenomenon, called acoustic levitation, has numerous applications in containerless study and processing of materials. For instance, acoustic levitation has been used in biochemical analysis,[1,2] and even small animals have been suspended.[3]

Although it is possible to levitate a sample by acoustic means for long periods of time, instabilities can appear under certain conditions.[4] A single-mode levitation device has a cavity, either open or closed, driven at one of its length resonances with very high sound pressure amplitudes. It has been shown that when a single-mode acoustic levitator is driven at a frequency slightly above the resonance of its empty cavity, oscillational instabilities can affect the levitated object. A cause of such instabilities is the change of the resonance frequency of the cavity due to the presence and motion of the object.[5] One of the motivations of the work described in this paper has been to further understand the causes of oscillational instabilities and the physical effects in acoustic levitation devices.

An investigation on the resonance frequency shift of a cylindrical cavity induced by a rigid sphere in its interior is presented. The position of the sphere can be any point along the axis of the cavity. Several studies considering the same problem are available in literature. An analytical equation to describe the resonance frequency shift for a plane-wave mode in a chamber containing a rigid sphere was first obtained by Leung *et al.*[6] The calculation of the frequency shift was based on a Green's function, considering the scattering of the acoustic wave on the sphere surface, but neglecting the interactions of the scattered wave on the cavity walls. The equation describes the resonance frequency shift to second order in $k_l R$, where k_l is the wave number corresponding to the resonance of the empty cavity and R is the radius of the sphere. A higher order approximation was, however, not possible.

Numerical calculations of the resonance frequency shift in finite cylindrical ducts produced by blockages of different shapes, including spheres, have been presented by El-Raheb and Wagner.[7] Their calculations are valid for both small and large samples with respect to the dimensions of the duct, and frequency shifts for longitudinal and transverse acoustic resonances can be calculated.

The study of the resonance frequency shift in a cylindrical cavity produced by a solid sphere was experimentally extended by Barmatz *et al.*[8] They analyzed the dependence of the frequency shift on the ratio between the diameter of the sphere and the height of the cavity and on the ratio of the cross-section areas between the sphere and the cavity.

An analytical expression for the resonance frequency shift of a plane-wave mode was obtained by studying the local perturbation of the acoustic field around a solid sphere by Curzon and Plant.[9] The expression was valid for very small spheres compared with the wavelength. If the terms depending on $k_l R$ in the equation given by Leung *et al.* are neglected, it becomes the same as the expression deduced by Curzon and Plant. The equation was also obtained by Rud-

a)Electronic mail: arturo.orozco@ccadet.unam.mx

nick and Barmatz,[5] who used a Green's function decomposed into two parts, one corresponding to the contribution of an empty cavity mode of which resonance frequency is close to a shifted resonance of the cavity with the sphere, and the other representing the contribution of the remaining empty cavity modes.

Following a similar procedure as the one reported by Leung *et al.*, a theory was developed by Mehl and Hill to obtain the eigenfrequencies of an arbitrary cavity shape with an internal sphere.[10] The change in the resonance frequencies of a cylindrical cavity produced by an eccentric sphere was theoretically studied by Roumeliotis *et al.*[11,12] They considered both rigid and pressure-release boundary conditions and also investigated the effect on rectangular cavities.[11,13] For the case of rigid boundaries, a plane-wave mode, and a solid sphere along the axis of the cavity, their equations reduce to the same result given by Curzon and Plant. Roumeliotis *et al.* also deduced an expression for the resonance frequency shifts in a spherical cavity with an eccentric sphere.[14]

Recently the resonance frequency shift produced by a solid sphere in an open cavity of a single-axis acoustic levitator has been investigated by Xie and Wei.[15] Based on the boundary element method (BEM), they presented numerical results, which coincide quite well with Leung's equation under the consideration of an equivalent volume for the cavity. However, Leung's equation predicted a lower value of the frequency shift compared to the numerical results when the sphere was located in a pressure maximum; this difference increased with the size of the sphere.

The resonance frequency shift of a plane-wave mode in a rectangular cavity induced by a spherical solid, with a diameter comparable to the length of the square cross section of the cavity, has been studied by Cordero and Mujica.[16] In this case, the main consideration was the change, due to the presence of the obstacle, of the cross-section area of the region occupied by the acoustic wave. Their theoretical calculations, obtained from a transcendental equation in terms of the confluent Heun function, are in good agreement with experimental results for small spheres compared with the wavelength.

In this paper it is shown that the Boltzmann–Ehrenfest principle of adiabatic invariance can be applied to predict the resonance frequency shift of a cylindrical cavity brought about by a solid sphere. The acoustic field in the cavity without the perturbation of the object corresponds to a plane-wave mode. An analytical expression is obtained to describe the resonance shift; the approximation in $k_l R$ is of the fourth order. It is proved by means of numerical calculations based on the BEM that this expression gives a better approach than Leung's equation. The theoretical results are also confirmed by experiments. It is illustrated that the Boltzmann–Ehrenfest principle is a powerful tool that allows the determination of the resonance frequency shift in a simpler and more direct way than the theoretical methods reported previously.

II. THEORY

A. The Boltzmann–Ehrenfest principle

Adiabatic invariants give approximate conservation laws for open systems.[17,18] A lossless system becomes open when it is affected by an external influence, and quantities such as the total energy are no longer conserved. Here the term *adiabatic* implies that any parameter associated with a change in the system is to be varied sufficiently slowly.

According to the Boltzmann–Ehrenfest principle, the ratio between the total energy of the sound field in a resonator (including an object) and the particular resonant frequency of the system is invariant provided the system is friction free and behaves linearly (i.e., its motion can be described by linear differential equations with amplitude-independent coefficients).[17,18] The Boltzmann–Ehrenfest principle has been successfully used to predict the torque acting on a cylinder suspended in a cylindrical chamber of an acoustic levitator by Putterman *et al.*[18]

Consider a closed cavity with a stationary sound wave in its interior. The frequency of the acoustic field corresponds to one of the resonance frequencies of the cavity. It should be noted that the system is assumed to be lossless; therefore, in order to have a finite value of the total acoustic energy, the excitation had to cease when the sound field was established.

Let E_o be the total acoustic energy of the stationary wave in the interior of the empty cavity at the resonance angular frequency ω_o. Now, let a sphere be expanded adiabatically with its center kept fixed in the interior of the cavity. In this process, the volume of the sphere is changed from zero to its final value, V_s. The work done on the system, W, during the total expansion is given by

$$W = \int_0^{V_s} f_n dV, \tag{1}$$

where f_n is the time-averaged acoustic normal force per unit area on the surface of the sphere.

Thus, the final total acoustic energy in the cavity will be $E_s = E_o + W$. Here the purpose is to find the new resonance frequency, ω_s, as a function of the size and position of a rigid sphere along the axis of a cylindrical cavity.

According to the Boltzmann–Ehrenfest principle,

$$E_o/\omega_o = E_s/\omega_s = (E_o + W)/\omega_s. \tag{2}$$

From here, we can obtain

$$(\omega_s - \omega_o)/\omega_o = W/E_o. \tag{3}$$

B. Normal acoustic force on a solid sphere

Consider a solid sphere in the interior of a cylindrical cavity with radius R_c and length L filled with an inviscid gas. Inside the cavity, with its walls assumed rigid, there is a stationary plane wave in resonance with a wave number k_l. Here it will be assumed that $k_l R \ll 1$, where R is the radius of the sphere; therefore, an acoustic resonance in the sphere is not possible. In addition, since the acoustic impedance of the sphere material is much larger than the impedance of the gas, the sphere can be considered rigid.

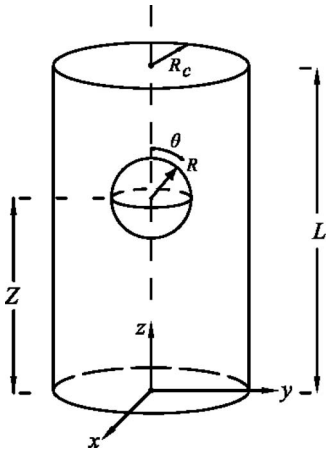


FIG. 1. Cylindrical cavity with the sphere in its axis and the reference system used.

Assume a cylindrical coordinate system as illustrated in Fig. 1; the position of the center of the sphere is Z. It will be considered that the sound pressure amplitude inside the cavity, p, is the sum of the sound pressure amplitude in a plane-wave mode of the empty cavity, $p_l = A\cos(k_l z)$, plus the sound pressure amplitude of the scattered acoustic wave, p_s. The time dependence of the acoustic variables is considered as $\exp(-i\omega t)$.

First, the acoustic pressure and particle velocity on the sphere surface will be determined. By using spherical coordinates with the origin at the center of the sphere, p_l can be expressed as

$$p_l = A\cos[k_l(r\cos\theta + Z)] = (A/2)(e^{ik_lZ}e^{ik_lr\cos\theta} + e^{-ik_lZ}e^{-ik_lr\cos\theta}). \tag{4}$$

Now, with the identity[19]

$$e^{ik_lr\cos\theta} = \sum_{n=0}^{\infty}(2n+1)i^n j_n(k_lr)P_n(\cos\theta), \tag{5}$$

where $j_n(k_lr)$ is the spherical Bessel function and $P_n(\cos\theta)$ is the Legendre polynomial, Eq. (4) yields

$$p_l = \sum_{n=0}^{\infty} B_n(2n+1)j_n(k_lr)P_n(\cos\theta), \tag{6}$$

where

$$B_n = (A/2)i^n[\exp(ik_lZ) + (-1)^n\exp(-ik_lZ)]. \tag{7}$$

It will be assumed that the sphere is small enough and that it is sufficiently far from the cavity walls that the scattered wave is due only to the presence of the sphere. Therefore, the interaction between the sphere and the walls is neglected. In this way, let the complex amplitude of the scattered acoustic wave be described by[6,20]

$$p_s = \sum_{n=0}^{\infty} C_n h_n^{(1)}(k_lr)P_n(\cos\theta), \tag{8}$$

where C_n is a constant to be determined and $h_n^{(1)}(k_lr)$ is the spherical Hankel function of the first kind.

Since the sphere is rigid, the component of the particle velocity normal to the surface of the sphere must be zero; therefore, by applying this condition

$$\left(\frac{\partial p_l}{\partial r}\right)_{r=R} = -\left(\frac{\partial p_s}{\partial r}\right)_{r=R}. \tag{9}$$

Substitution of Eqs. (6) and (8) into Eq. (9) leads to the values of C_n as

$$C_n = -B_n\frac{(2n+1)j_n'(k_lR)}{h_n^{(1)'}(k_lR)}. \tag{10}$$

As a result, the total sound pressure amplitude at a point on the sphere surface is given by

$$p(R,\theta) = \sum_{n=0}^{\infty}(2n+1)B_nP_n(\cos\theta) \times\left[j_n(k_lR) - \frac{j_n'(k_lR)}{h_n^{(1)'}(k_lR)}h_n^{(1)}(k_lR)\right]. \tag{11}$$

In the limit of small k_lR, Eq. (11) can be approximated at the order of $(k_lR)^2$; in this way, the first three terms of the summation are conserved. In addition, by substituting the spherical Bessel and Hankel functions in terms of sine and cosine functions, and by expressing these trigonometric relations as series of k_lR, after lengthy but straightforward calculations, Eq. (11) reduces to

$$p(R,\theta) = A\cos(k_lZ)\left[1 - \tfrac{1}{2}(k_lR)^2\right] - A\tfrac{3}{2}k_lR\sin(k_lZ)P_1(\cos\theta) - A\tfrac{5}{9}(k_lR)^2\cos(k_lZ)P_2(\cos\theta). \tag{12}$$

An inviscid fluid and an immovable sphere have been considered; moreover, taking into account that the system is independent of the azimuthal angle, it follows that the fluid on the surface of the solid will oscillate only in the θ direction. Thus, the component in that direction of the complex particle velocity amplitude is determined by means of Euler's equation as $u_\theta = (i\omega\rho r)^{-1}(\partial p/\partial\theta)$, where ρ is the mean density of the air in the cavity. Use of Eq. (12) yields

$$u_\theta(R,\theta) = -i\frac{3}{2}\frac{A}{\rho c}\sin(k_lZ)\sin\theta - i\frac{5}{3}\frac{A}{\rho c}k_lR\cos(k_lZ)\cos\theta\sin\theta, \tag{13}$$

where c is the speed of sound in air.

The time-averaged acoustic normal force per unit area on the sphere is given by[20]

$$f_n = \frac{1}{2\rho c^2}\langle p^2\rangle - \frac{1}{2}\rho\langle \boldsymbol{u}\cdot\boldsymbol{u}\rangle, \tag{14}$$

where $\langle\cdot\rangle$ indicates the time average and p and $\boldsymbol{u}$ are, respectively, the sound pressure and the particle velocity evaluated on the surface of the sphere. Substitution of Eqs. (12) and (13) into Eq. (14) leads to

$$f_n(R,\theta)=\frac{A^2}{4\rho c^2}\left\{\left[\cos^2(k_lZ)-\frac{9}{4}\sin^2(k_lZ)\sin^2\theta\right]-\cos(k_lZ)\sin(k_lZ)\cos\theta[3+5\sin^2\theta]k_lR\right.$$
$$+\left[\frac{9}{4}\sin^2(k_lZ)\cos^2\theta+\cos^2(k_lZ)\left(\frac{1}{9}-\frac{5}{3}\cos^2\theta-\frac{25}{9}\cos^2\theta\sin^2\theta\right)\right](k_lR)^2$$
$$\left.+\cos(k_lZ)\sin(k_lZ)\left[-\frac{1}{6}\cos\theta+\frac{5}{2}\cos^3\theta\right](k_lR)^3+\cos^2(k_lZ)\left[\frac{1}{324}-\frac{5}{54}\cos^2\theta+\frac{25}{36}\cos^4\theta\right](k_lR)^4\right\}. \quad (15)$$

C. Determination of the work

The total normal acoustic force exerted on the surface of the sphere, F_n, can now be evaluated as

$$F_n=\int_S f_n(R,\theta)da, \quad (16)$$

where S is the total surface of the sphere. Use of Eq. (15) gives

$$F_n=\frac{A^2\pi R^2}{2\rho c^2}\left\{-\frac{1}{2}+\frac{5}{2}\cos(2k_lZ)\right.$$
$$+\left[-\frac{7}{108}-\frac{169}{108}\cos(2k_lZ)\right](k_lR)^2$$
$$\left.+\left[\frac{1}{9}+\frac{1}{9}\cos(2k_lZ)\right](k_lR)^4\right\}. \quad (17)$$

The work can now be determined with Eq. (1) using F_n as

$$W=\int_0^R F_n dr. \quad (18)$$

Substitution of Eq. (17) into Eq. (18) produces

$$W=\frac{A^2}{4\rho c^2}V_s\left\{\left[-\frac{1}{4}+\frac{5}{4}\cos(2k_lZ)\right]\right.$$
$$+\left[-\frac{7}{360}-\frac{169}{360}\cos(2k_lZ)\right](k_lR)^2$$
$$\left.+\left[\frac{1}{42}+\frac{1}{42}\cos(2k_lZ)\right](k_lR)^4\right\}. \quad (19)$$

D. Resonance frequency shift

The total acoustic energy for a plane-wave mode in the cylindrical cavity without the sphere is

$$E_o=\int_{V_c}\left[\frac{1}{2\rho c^2}\langle p(z)^2\rangle+\frac{1}{2}\rho\langle \boldsymbol{u}(z)\cdot\boldsymbol{u}(z)\rangle\right]dV. \quad (20)$$

With $p(z)=A\cos(k_lz)\exp(-i\omega t)$ and the use of Euler's equation to calculate $\boldsymbol{u}(z)$, Eq. (20) gives $E_o=V_cA^2/(4\rho c^2)$. It should be taken into account that $k_l=l\pi/L$, where l is a natural number. Finally, the use of this result and Eq. (19) allows the calculation of the resonance frequency shift for the cylindrical cavity produced by a small solid sphere in its interior according to Eq. (3):

$$\frac{\omega_s-\omega_l}{\omega_l}=\frac{V_s}{V_c}\left\{\left[-\frac{1}{4}+\frac{5}{4}\cos(2k_lZ)\right]\right.$$
$$-\left[\frac{7}{360}+\frac{169}{360}\cos(2k_lZ)\right](k_lR)^2$$
$$\left.+\left[\frac{1}{42}+\frac{1}{42}\cos(2k_lZ)\right](k_lR)^4\right\}, \quad (21)$$

where ω_l is the angular frequency of the plane-wave mode l in the empty cylindrical cavity.

E. Comparison with published results

If only the k_lR-independent terms are kept in Eq. (21), it reduces to the same expression obtained previously by other authors,[5,9] which is valid for a plane-wave mode in a cavity with rigid objects significantly smaller than the wavelength and not necessarily spherical.

The corresponding equation for the resonance frequency shift determined by Leung *et al.* is

$$\frac{\omega_s-\omega_o}{\omega_o}=\frac{V_s}{V_c}\left\{\left[-\frac{1}{4}+\frac{5}{4}\cos(2k_lZ)\right]\right.$$
$$\left.-\left[\frac{67}{360}+\frac{229}{360}\cos(2k_lZ)\right](k_lR)^2\right\}. \quad (22)$$

It can be seen that the coefficient of the term $(k_lR)^2$ in Eq. (21) is slightly different from Leung's equation. When $\cos(2k_lZ)=-1$, which correspond to the minimum of the resonance frequency shift, the value of this coefficient is the same in both cases; in fact, Eqs. (21) and (22) give the same result since the coefficient of the term $(k_lR)^4$ in Eq. (21) becomes equal to zero at that position of the sample. If the size of the sphere is very small compared with the wavelength, the difference between the equation obtained by Leung *et al.* and Eq. (21) is negligible.

However, if k_lR tends to unity, the term involving $(k_lR)^2$ is not negligible. Moreover, the term with the factor $(k_lR)^4$ produces a better approach. It is shown below that Eq. (21) gives a more general approximation than Eq. (22) in that case. In addition, the procedure followed by Leung *et al.* allowed a description only to second order in k_lR; complex terms were encountered with a higher order approach. This restriction did not appear by using the Boltzmann–Ehrenfest principle.

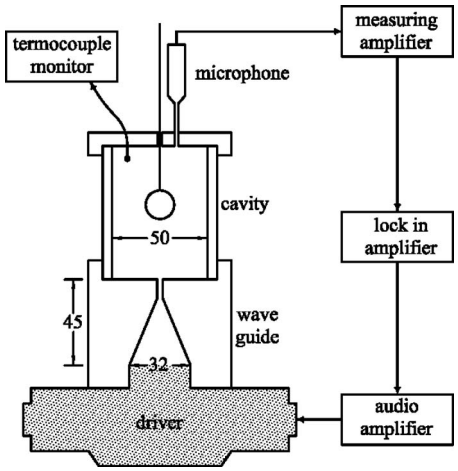


FIG. 2. Diagram of the experimental setup used. The indicated dimensions are in millimeters.

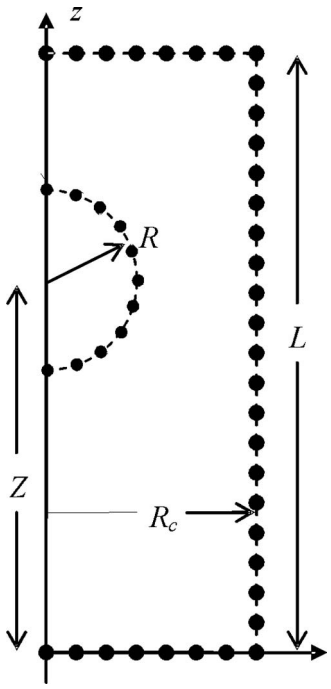


FIG. 3. Example of the axisymmetrical BEM mesh used in the calculations.

III. EXPERIMENT

The measurements of the resonance frequency shift were carried out using the experimental setup illustrated in Fig. 2. The cylindrical cavity was formed using an acrylic tube with an inner radius of 2.54 cm and a wall thickness of 0.6 cm; several lengths of the cavity were used. The acoustic wave was produced by means of a driver from a horn loudspeaker. To couple the driver to the cylindrical cavity, a waveguide made of polyvinyl chloride (PVC), a widely used thermoplastic, was employed. The opening of the waveguide into the cavity had an inner diameter of 0.3 cm and was centered on the axis of the cylinder. The top of the cavity was also made of PVC, and it had a thickness of 1.0 cm. Several spherical samples, made of glass, were used. They were suspended along the axis of the cylinder by means of a thin optic fiber 0.25 mm in diameter attached to them.

The sound pressure amplitude was measured on the top of the cavity by using a probe microphone (B&K 4182), which has an outer diameter of 1.24 mm and was located 1.25 cm from the axis of the cylinder. The resonance frequency of the cavity was determined from the curve of the frequency response obtained with a lock-in amplifier (SR 8500). In this case, a sinusoidal sweep signal generated by the lock-in amplifier, passing through an audio amplifier, was used as the signal to excite the driver.

The main source of error in the experiment was the variation of the temperature, which changed the sound speed and, as a consequence, the resonance frequency. For that reason, the temperature was registered with a thermocouple, type E, connected to a thermocouple monitor (SR 360).

In order to reduce the heat delivered by the driver, the amplitude of the voltage to excite that transducer was required to have a relatively low value. The floor noise in the empty cavity was first measured, which was less than 25 dB in all cases. Taking into account the measured value of the floor noise, the selected voltage was adjusted to produce about 95 dB at resonance at the top of the empty cavity. For this value, it was observed that the heat delivered by the voice coil of the driver increased the temperature inside the cavity sufficiently slow; in addition, the effect of the floor noise was negligible.

Measurements were made of the resonance frequency as a function of the position of a given sphere along the axis of the cavity. In each set of measurements, the initial position was close to the opening of the waveguide into the cavity; the position of the sphere was changed until it was close to the top of the cavity, and the measurements were repeated from the top to the initial position. In this way two measurements were made at each position, and the mean value is reported.

Due to the variations in temperature, the resonance frequency of the empty cavity was measured at the beginning and at the end of each set of measurements of the resonance frequency shift. The mean value of these two measurements was considered as the resonance frequency of the empty cavity.

IV. NUMERICAL SIMULATIONS

The resonance frequencies of interest for the various configurations were calculated numerically using the BEM. Since the setup is independent of the azimuthal angle, a BEM formulation for axisymmetrical bodies was used, where only the generator of the structure is meshed into one-dimensional elements,[21] as shown in Fig. 3.

As a rule of thumb, it is required that a BEM mesh contains at least six elements per wavelength. In the simulations for this paper, 24 quadratic elements were used along the generator of the cylindrical cavity, distributed with 20 elements on the side and 7 on each lid. The spherical object was modeled using ten elements in all cases. Considering the dimensions and frequencies employed, this mesh is more

than sufficient. However, as a check, some resonance frequency calculations were repeated with rougher meshes, yielding negligible differences.

The results were further verified using the analytical solution for the eigenfrequencies in an empty cylindrical cavity. The numerically calculated eigenfrequency never fell more than 1 Hz apart from the theoretical value.

The BEM, unlike the finite element method (FEM), does not readily provide direct computation of the eigenfrequencies. In the FEM, the homogeneous differential wave equation is used for this purpose. The BEM, on the other hand, shows the eigenfrequencies as maxima, as a function of the frequency, of the condition numbers of its coefficient matrices. In the present calculation, an iterative process has been employed with the purpose of finding these maxima, with successive BEM calculations at varying frequency. If this iteration process is properly adjusted, the BEM will give eigenfrequency values as accurate as FEM can, given the same mesh density. There were two reasons for using BEM instead of FEM in this investigation: (i) the geometrical modifications for moving or scaling the small sphere are simple to do automatically with the discretized BEM model, whereas a FEM model would require the geometry to be remeshed for each configuration, and (ii) the BEM code was already available; it has been developed for research purposes by a group that includes one of the authors.

In this paper, the eigenfrequencies of the cavity-sphere setup are calculated as a function of the sphere position, the sphere radius, and the cavity radius. In the first case, calculations were made in steps of $0.005/L$, where L is the cylinder length. In the second and third cases, the sphere radius and the cavity radius were increased in steps of 0.02 cm. The number of values is large enough to allow showing the results in the plots as continuous lines. The time needed for the calculation was of about 3 min per eigenfrequency in a modern desktop computer, giving a few hours to produce any of the plots presented later in this paper.

V. RESULTS AND ANALYSIS

A. Position of the sphere

According to published papers,[6,8] the difference between experimental values of the resonance frequency shift and the analytical results obtained with the expression deduced by Leung *et al.* increases with the order of the mode. Here results for the second and third plane-wave modes are presented since they are more illustrative.

The difference between Eqs. (21) and (22) tends to be negligible as k_lR approaches zero, but it becomes noticeable for values of k_lR close to unity. An example of the resonance frequency shift as a function of the position of a sphere along the axis of the cylindrical cavity is shown in Fig. 4. In this case the analyzed resonance corresponds to the second plane-wave mode of a cavity with $L=7.05$ cm; the average resonance frequency measured in the empty cavity was 4867.0 Hz. A relatively large sphere with a radius $R=1.47$ cm was used in the experiments, for which $k_2R=1.31$. The resonance frequency shift obtained with Eq. (21), which gives an approximation to the fourth order in k_2R,

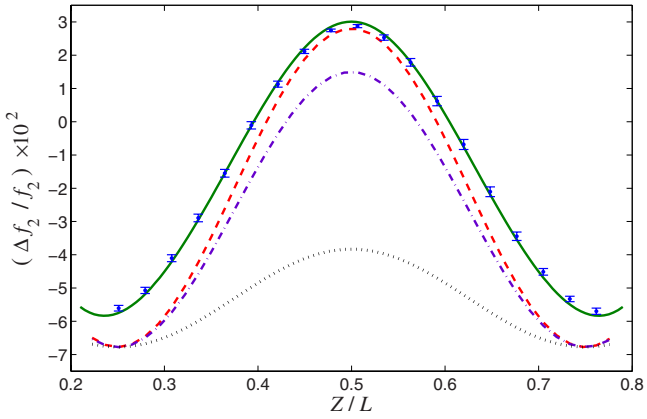


FIG. 4. (Color online) Dependence of the second plane-wave mode frequency shift on the position of a sphere along the axis of the cavity. Numerical simulations (—), experimental data (•) with estimated uncertainties, results based on the Boltzmann–Ehrenfest principle [Eq. (21)] approximated to $(k_2R)^4$ (– –) and to $(k_2R)^2$ (·–·), and results predicted with Leung's equation (···). Here $k_2R=1.31$.

compares well with the experimental data despite the fact that k_2R is larger than unity. The difference between the predictions of Eq. (21) and the experimental results decreases around the maximum of the resonance frequency shift. An uncertainty analysis of the experimental data using Eq. (21) as measurement model has been performed,[22] and the result is included as error bars in Fig. 4.

If the terms with $(k_2R)^4$ are neglected in Eq. (21), the resulting expression can be compared more appropriately to Eq. (22) (Leung's equation) since both are approximated to the second order in k_2R. As can be seen in Fig. 4, the results from Eq. (21) approximated to second order in k_2R and from Leung's equation are the same in the minima of the resonance frequency shift; however, the difference between the results of the two equations increases around the maximum of the frequency shift. The prediction obtained from the Boltzmann–Ehrenfest principle gives a better approximation than Leung's equation for large spheres.

The results from the numerical simulations based on the BEM are also depicted in Fig. 4; they agree very well with the experimental data. During the measurement, the temperature inside the cavity increased in 0.5 °C, which gave a variation in the resonance frequency of the empty cavity of approximately 4 Hz. The maximum sound pressure level measured at resonance was 96.7 dB (referred to 20 μPa) and the minimum value at resonance was 92.9 dB.

Another example of the resonance frequency shift as a function of the position of the sphere is shown in Fig. 5, where the studied mode corresponded to $l=3$. It was decided to use a larger cavity and not a smaller wavelength in order to avoid the effect of any resonance of the system driver-waveguide in the measurements of the resonance frequency shift. In this way, a cylindrical cavity with $L=10.57$ cm was employed, the third mode of which was expected to have a resonance frequency similar to the one corresponding to the second mode of the previous case of Fig. 4. The average experimental resonance frequency of the empty cavity was equal to 4838.1 Hz. The difference with the expected value of 4867 Hz is explained by the fact that the temperature in

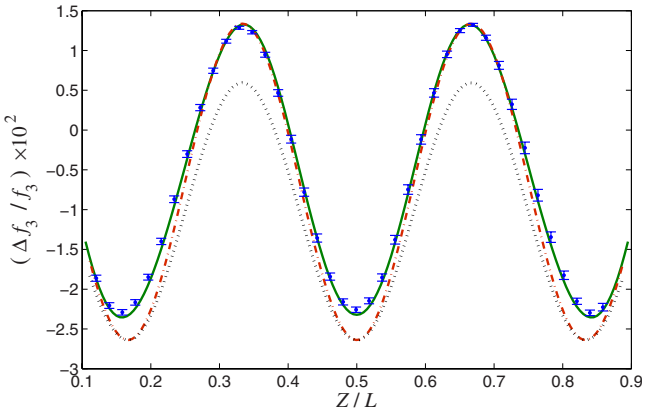


FIG. 5. (Color online) Relative resonance frequency shift for the third plane-wave mode of a cylindrical cavity as a function of the position of a sphere along the axis. Experiment (•), numerical simulations (—), results based on the Boltzmann–Ehrenfest principle (-·-), and predictions with Leung's equation (···). Here, the length of the cavity was $L=10.57$ cm and $k_3R=0.96$. Estimated uncertainties for the experimental data are also shown.

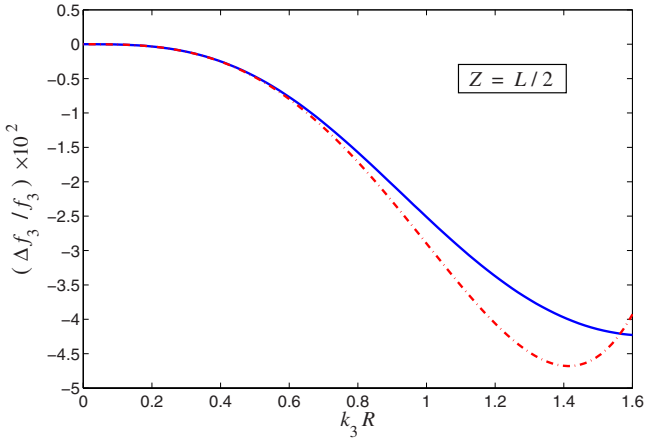


FIG. 6. (Color online) Dependence of the resonance frequency shift for the third plane-wave mode on the size of the sphere, which is assumed to be located in the center of the cylindrical cavity with length $L=10.57$ cm. Numerical simulations based on the BEM (—). In this case, Eq. (21) gives the same curve (-·-) as Eq. (22).

the laboratory was lower when measuring the third mode than it was when the experiment for the second mode was carried out.

The curves in Fig. 5 correspond to a sphere with a radius of 1.075 cm, which gives a value of $k_3R=0.96$. A very good agreement between the experimental data and the results of the numerical simulations can be observed again. For the conditions of this experiment, the contribution of the terms with the factor $(k_lR)^4$ in Eq. (21) is very small; therefore, the curve in Fig. 5 obtained with Eq. (21) is an approximation to second order in k_lR. This curve compares very well with the experimental data around the maxima of the relative resonance frequency shift. However, the difference between this curve and the measurements increases in the neighborhood of the minima. It can also be observed that Eq. (21) gives a better approximation to the experimental results than Leung's equation.

In the experiment, the maximum value of the sound pressure measured with the microphone at resonance was 92.5 dB and the minimum value was 89.4 dB. The temperature in the cavity increased 0.8 °C from the beginning to the end of the set of measurements, which gave a variation of approximately 7 Hz in the resonance frequency with the cavity empty.

Measurements of the resonance frequency were quite repeatable, within a fraction of 1 Hz, provided that the mean temperature remained constant. This behavior has already been reported in literature.[8] The numerical results show a very small but systematic overestimation of the resonance displacement when compared with the experimental data. This overestimation might be due to factors not taken into account in the numerical simulations, mainly viscosity, thermal conduction, and the effect of the opening in the cavity connecting it to the waveguide.

B. Size of the sphere

The effect of the size of the sphere on the resonance frequency shift was also analyzed. The result for the mode $l=3$ with a sphere centered in the cylindrical cavity, 10.57 cm long, can be seen in Fig. 6, where numerical simulations based on the BEM are compared with the predictions obtained from Eq. (21). As pointed out above, Eqs. (21) and (22) give the same result when the sphere is located in a position corresponding to a pressure node of a pure plane-wave mode, and the resonance frequency shift is negative. According to the curves in Fig. 6, the analytical results compare in an excellent way to the numerical simulations for values of k_3R smaller than 0.6. The difference between the two curves increases with k_3R until it reaches its maximum for the value of 1.4, where the curve representing the analytical equation has a minimum. The two curves intersect each other at approximately $k_3R=1.6$, and they start diverging very rapidly from that point. Practically the same graph was obtained with the third mode and a cavity 7.05 cm long; in that case, the difference between the numerical and the analytical curves increases with the radius of the sphere and the two curves also intersect to each other at approximately $k_3R=1.6$.

When the center of the sphere is located in the position corresponding to a pressure antinode of a plane-wave mode of an empty cylindrical cavity, positive resonance frequency shifts are produced. As an example, the results for the mode $l=3$ in a cavity with a length $L=10.57$ cm and $Z=L/3$ are shown in Fig. 7 as a function of the size of the sphere. In this figure, the values of the resonance frequency shift obtained by means of Eq. (21), taking into account the term with the factor $(k_3R)^4$, are a better approach to the numerical simulations for large spheres, up to approximately $k_3R=1.2$. In order to show the contribution of the term with the factor $(k_3R)^4$, this term is neglected in Eq. (21) to get an approximation of that equation to the second order in k_3R. A good agreement is obtained between the numerical simulations and the second order approximation in k_3R of Eq. (21) up to about $k_3R=1$. On the other hand, the results given by Leung's equation become significantly different from the numerical simulations from approximately $k_3R=0.8$.

As can be observed, Leung's equation provides a very good accuracy for values of k_lR smaller than 0.6 for any

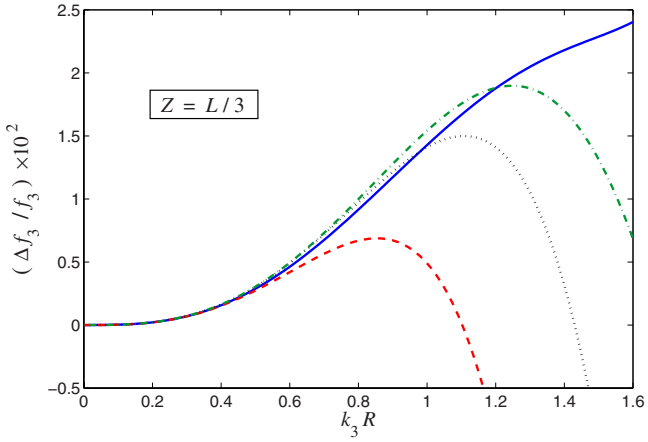


FIG. 7. (Color online) Relative resonance frequency shift of the $l=3$ mode of a cylindrical cavity as a function of the size of a sphere in its interior. The length of the cavity is $L=10.57$ cm, and the center of the sphere is assumed to be on the axis at $Z=L/3$. Numerical simulations based on the BEM (—), Leung's equation (– –), results based on the Boltzmann–Ehrenfest principle [Eq. (21)] approximated to $(k_3R)^2$ (···) and to $(k_3R)^4$ (–·–).

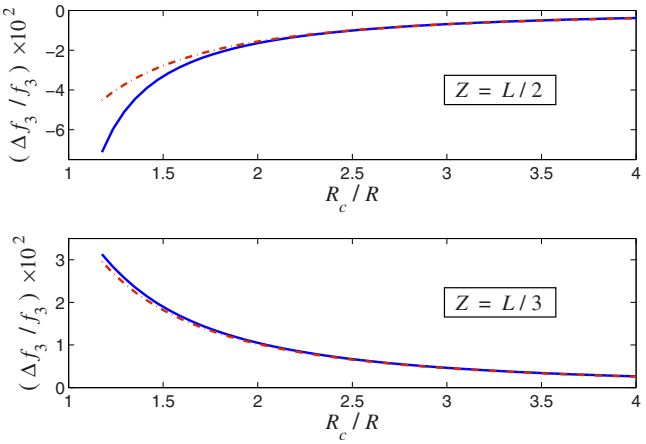


FIG. 8. (Color online) Relative resonance frequency shift for the mode $l=3$ as a function of the radius of the cavity. The center of the sphere, with a radius $R=0.34$ cm ($k_3R=0.3$), has been assumed to be located at $Z=L/2$ for the upper graph and at $Z=L/3$ for the lower graph. The curves correspond to analytical solutions given by Eq. (21) (·–·) and to numerical simulations based on the BEM (—).

position of a sphere along the axis of the cylindrical cavity. When a sphere is located with its center in the corresponding position of a pressure node of the empty cavity, a good agreement is achieved for larger values of k_lR. However, when the sphere is located in a pressure antinode of the empty cavity, the approximation given by this equation is rapidly reduced as the value of k_lR increases. This fact was also explicitly pointed out by Xie and Wei.[15] They carried out simulations with a sphere located in the middle of an open cavity assuming an equivalent closed volume. Their results showed that the simulation agreed well with Leung's theory in the first and third modes, where the position of the sphere coincided with the sound pressure node; in contrast, a noticeable difference appeared at the second resonance for a large radius of the sphere, in this case placed at the antinode.

C. Radius of the cavity

To observe the effect of the lateral walls of the cavity on the resonance frequency shift, a small sphere was assumed to be fixed at a certain position Z in the axis of a cylindrical cavity; the resonance frequency shift was calculated for different values of the radius of the chamber. Figure 8 shows the results corresponding to the third plane wave mode of the empty cavity with a length $L=10.57$ cm and a sphere with a radius $R=0.34$ cm. The sound speed was considered to be $c=342.3$ m/s, which gives $f_3=4857.6$ Hz and $k_3R=0.3$. The relative resonance frequency shifts are depicted for two positions of the center of the sphere, for $Z=L/2$ (at the pressure node of the empty cavity) and for $Z=L/3$ (the location of the pressure antinode). Numerical simulations obtained by means of the BEM are compared with the analytical results given by Eq. (21).

According to the results in Fig. 8, when $Z=L/3$, the discrepancy between the numerical simulations and the analytical results is quite small even when the radius of the sphere is similar to the radius of the cavity. On the other hand, for $Z=L/2$, the differences between the numerical and the analytical curves become noticeable for $R_c/R<1.7$ and rapidly increases as R_c/R approaches to unity. It appears that the effect of the scattering of the sound wave by the walls of the cavity is higher when the sphere is located in a velocity antinode than when the solid is located in a pressure antinode. Thus, the perturbation of the fluid motion caused by the presence of the sphere probably creates more interaction with the walls than the perturbation of the sound pressure, explaining the poorer approximation to the actual frequency shift of all analytical expressions, which assume no such interaction.

It should be observed that there is a pressure antinode on each of the circular walls of the empty cylindrical cavity when a plane-wave mode exists. Therefore, when the sphere is located close to a circular wall of the cavity, the effect of the scattering of the sound wave produced by the boundaries is expected to be small. This can be observed in Fig. 5, where the numerical simulations of the resonance frequency shift agree quite well with the curve drawn based on Eq. (21) for positions of the sphere very close to the circular walls.

The results of Fig. 8 are consistent with the observations of Cordero and Mujica.[16] They used a sphere 6.35 mm in diameter within a square-section cavity, 6.8 mm in side and 100.0 mm long. It was reported that the differences between measurements of the resonance frequency shift and theoretical results given by Leung's equation were larger at the positions of the velocity antinodes of a pure plane-wave mode than when the sphere was placed at the pressure antinodes.

In the experiments of Cordero and Mujica, the sphere to cavity cross-sectional area ratio was equal to 1.46, which corresponds to the value $R_c/R=1.21$ in Fig. 8. Considering this relation, the experimental values of the resonance frequency shift obtained by Cordero and Mujica for the third plane-wave mode are similar to the numerical results of Fig. 8, which gives a further verification of the numerical results.

VI. CONCLUSIONS

Based on the results presented in this paper, it can be concluded that the Boltzmann–Ehrenfest principle allowed

the determination of an analytical expression to predict the resonance frequency shift of a plane-wave mode in a cylindrical cavity caused by a solid sphere located at any position along the axis of the cavity. This method is more direct and simpler than a previously reported one based on a Green's function. In contrast with the latter, it was possible to predict the resonance frequency shift to the fourth order in $k_l R$. Furthermore, the obtained expression was shown to be more general than the one previously published; the equation based on the Boltzmann–Ehrenfest principle agreed better with experimental and numerical results for large spheres, even for values of $k_l R > 1$.

In addition, it has been shown that the effect of scattering of the sound waves by the boundaries of the cavity on the resonance frequency shift, when the radius of the cylindrical cavity is less than twice the radius of the sphere, is more significant when the solid is located at the velocity antinode of the pure plane-wave mode than when the solid is placed at the pressure antinode.

ACKNOWLEDGMENTS

The authors wish to thank Professor Seth Putterman of UCLA for his suggestion on the use of the Boltzmann–Ehrenfest principle. The support from DGAPA-UNAM, Grant No. IN-114407, is also acknowledged.

[1]M. Wiklund, J. Toivonen, M. Tirri, P. Hänninen, and H. M. Hertz, "Ultrasonic enrichment of microspheres for ultrasensitive biomedical analysis in confocal laser-scanning fluorescence detection," J. Appl. Phys. **96**, 1242–1248 (2004).

[2]S. Santesson and S. Nilsson, "Airborne chemistry: Acoustic levitation in chemical analysis," Anal. Bioanal. Chem. **378**, 1705–1709 (2004).

[3]W. J. Xie, C. D. Cao, Y. J. Lü, Z. Y. Hong, and B. Wei, "Acoustic method for levitation of small living animals," Appl. Phys. Lett. **89**, 214102 (2006).

[4]R. C. Field and S. Scheeline, "Design and implementation of an efficient acoustically levitated drop reactor for *in stillo* measurements," Rev. Sci. Instrum. **78**, 125102 (2007).

[5]J. Rudnick and M. Barmatz, "Oscillational instabilities in single-mode acoustic levitator," J. Acoust. Soc. Am. **87**, 81–92 (1990).

[6]E. Leung, C. P. Lee, N. Jacobi, and T. G. Wang, "Resonance frequency shift of an acoustic chamber containing a rigid sphere," J. Acoust. Soc. Am. **72**, 615–620 (1982).

[7]M. El-Raheb and P. Wagner, "Acoustic propagation in rigid ducts with blockage," J. Acoust. Soc. Am. **72**, 1046–1055 (1982).

[8]M. Barmatz, J. L. Allen, and M. Gaspar, "Experimental investigation of the scattering effects of a sphere in a cylindrical resonant chamber," J. Acoust. Soc. Am. **73**, 725–732 (1983).

[9]F. L. Curzon and D. Plant, "Using perturbed resonant frequencies to study eigenmodes of an acoustic resonator," Am. J. Phys. **54**, 367–372 (1986).

[10]B. Mehl and R. N. Hill, "Acoustic eigenfrequencies of cavities with an internal obstacle: A modified perturbation theory," J. Acoust. Soc. Am. **85**, 1841–1851 (1989).

[11]J. A. Roumeliotis, J. D. Kanellopoulos, and J. G. Fikioris, "Acoustic eigenfrequencies of a cylindrical/rectangular cavity with an eccentric inner small sphere," J. Franklin Inst. **329**, 413–427 (1992).

[12]J. A. Roumeliotis, J. D. Kanellopoulos, and J. G. Fikioris, "Corrigendum," J. Franklin Inst. **330**, 1235–1236 (1993).

[13]J. A. Roumeliotis, "Eigenfrequencies of an acoustic rectangular cavity containing a rigid small sphere," J. Acoust. Soc. Am. **93**, 1710–1715 (1993).

[14]J. A. Roumeliotis, J. D. Kanellopoulos, and J. G. Fikioris, "Acoustic resonant frequency shifts is a spherical cavity with an eccentric inner small sphere," J. Acoust. Soc. Am. **90**, 1144–1148 (1991).

[15]W. J. Xie and B. Wei, "Resonance shift of single-axis acoustic levitation," Chin. Phys. Lett. **24**, 135–138 (2007).

[16]M. L. Cordero and N. Mujica, "Resonant frequency shifts induced by a large spherical object in an air-filled acoustic cavity," J. Acoust. Soc. Am. **121**, EL244–EL250 (2007).

[17]W. E. Smith, "Generalization of the Boltzmann-Ehrenfest adiabatic theorem in acoustics," J. Acoust. Soc. Am. **50**, 386–388 (1971).

[18]S. Putterman, J. Rudnick, and M. Barmatz, "Acoustic levitation and the Boltzmann-Ehrenfest principle," J. Acoust. Soc. Am. **85**, 68–71 (1989).

[19]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. II, Eq. (11.3.45).

[20]T. G. Wang and C. P. Lee, "Radiation pressure and acoustic levitation," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, 1998), Chap. 6, pp. 177–205.

[21]P. M. Juhl, "An axisymmetric integral equation formulation for free space non-axisymmetric radiation and scattering of a known incident wave," J. Sound Vib. **163**, 397–406 (1993).

[22]Guide to the Expression of Uncertainty in Measurement, BIPM, IEC, IFCC, ISO, IUPAP, IUPAC, OIML (1995).

A theoretical treatment of the long range propagation of impulsive signals under strongly ducted nocturnal conditions

Roger Waxler,[a] Kenneth E. Gilbert, and Carrick Talmadge
National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

(Received 26 March 2008; revised 31 July 2008; accepted 2 August 2008)

On clear nights, over flat land, a sound duct develops in which sound can carry to great distances. As is the case with all ducted propagation, there is strong dispersion so that a broadband signal undergoes severe distortion as it propagates. The signal received at long ranges from an impulsive source is a wave train, of much greater duration than the initial impulse, consisting of a series of arrivals. The long range ground to ground propagation of an impulsive signal in a typical nocturnal duct is studied and the natures of the various arrivals are explained. A direct connection is drawn between the meteorological and ground conditions and the structure of the propagated signal.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2980520]

PACS number(s): 43.28.Fp, 43.50.Pn, 43.50.Vt [VEO] Pages: 2742–2754

I. INTRODUCTION

The concern in this paper is with the long range near-ground to near-ground propagation of impulsive sound over flat land on clear and dry nights. On clear and dry nights there is generally a temperature inversion at the ground in which the mean temperature increases with increasing altitude.[1,2] The inversion typically extends up to altitudes of a few hundreds of meters. Above the inversion there is generally a strong geostrophic wind known as the nocturnal jet.[1–3] In this paper a theory is developed for the propagation of impulsive sound in the sound duct produced by such temperature inversions and winds. Upwind propagation will not be considered here, only the strongly ducted propagation directions will be studied.

Under the above conditions the atmospheric boundary layer is said to be stable since, in the inversion, turbulent mixing is suppressed.[1,2] Near the ground, atmospheric variability is expected to be small. There is variability driven, at lower altitudes, by internal waves which can be excited in the inversion[4] and, at higher altitudes, by the nocturnal jet which is accompanied by wind-shear driven turbulence. In this paper, however, the atmosphere will be assumed to be static and stratified. In particular, the effect of internal waves in the inversion, and turbulence above, will not be considered.

The work to be presented here may be considered the broadband complement to the narrowband structure discussed in Ref. 5. In that paper it was shown that, in typical nocturnal ducts at long ranges from a monotone source, the sound field near the ground is a superposition of a ground-coupled surface mode with a sum of airborne modes. The airborne modes will be referred to in this work as the refracted modes (rather than the less descriptive term "higher modes" used in Ref. 5).

In Refs. 6–8 it was noted that an impulse, propagated in typical nocturnal ducts to ranges of a few kilometers from its source, has a characteristic structure. The initial part of the propagated pulse consists of one or more distinct arrivals whose form is given qualitatively by ray theory. The frequency band of these initial arrivals is limited primarily by atmospheric attenuation so that, at ranges of a few kilometers, these initial arrivals can carry energy up to 1 kHz.[9–11] The propagated pulse ends with a low frequency tail with frequency content at or below 100 Hz. In a recent paper experimental and theoretical evidence has been reported, which identifies the low frequency tail with the ground-coupled surface mode.[5,12]

In Fig. 1 a typical pulse form, measured during a series of experiments performed at night over farmland in the Mississippi Delta, is shown. Figure 1(a) shows the pulse measured 20 m from the source and Fig. 1(b) shows the pulse obtained 1.6 km downwind from the source by a microphone near the ground. The structure described above is clearly visible. There is a distinct first arrival and low frequency tail. Note that the first arrival resembles an attenuated version of the initial pulse. The tail has undergone significant dispersion and bears no resemblance at all to the initial pulse.

It is the purpose of this paper to elucidate the nature of the various arrivals from an impulsive signal propagating to long ranges in a nocturnal duct. The effect of the ground surface, the temperature inversion, and the winds on the arrival structure will be discussed. Additionally, the theory and vocabulary required for more detailed analysis of experiments will be developed. The primary analytical tool employed will be the expansion in vertical modes developed in Ref. 13 for propagation in nocturnal ducts over a locally reacting ground surface. The arrival structure predicted by the modal expansion will be compared to the predictions of ray theory.

The organization of this paper is as follows. First the modal expansion is presented, then a discussion and presentation of the meteorological and ground impedance models

[a]Electronic mail: rwax@olemiss.edu

 © 2008 Acoustical Society of America

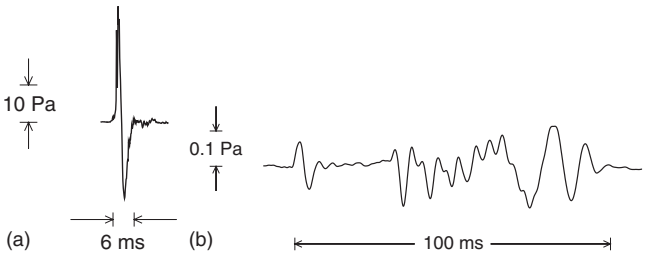


FIG. 1. An example pulse received 1.7 km from an impulsive source. (a) Signal measured 20 m from the source. (b) Signal measured 1.7 km from the source.

to be used follows. Propagation in the absence of wind is considered first and a characteristic arrival structure is found to be present at all ranges. Downwind propagation is considered next. The nocturnal jet is found to produce convergence zones in which the arrival structure alters significantly.

II. MODAL EXPANSION

The modal expansion developed in Ref. 13 is valid for long range, near-ground to near-ground, propagation over an impedance plane, under strongly ducted conditions. Three physical assumptions are made. Propagation angles are assumed to be small enough that the influence of the wind can be modeled by the effective sound speed approximation. In this approximation the effect of the wind on the propagation is modeled by adding the wind speed in the direction of propagation to the sound speed.[14–17] Propagation ranges are assumed to be long enough (in practice greater than a few hundred meters) so that the contribution from the continuous spectrum can be ignored. Propagation ranges are assumed to be short enough (in practice less than about 8 km) so that the upward refraction at high altitudes can be ignored.[18]

When the upward refraction at high altitudes can be ignored, the asymptotically constant approximation used in Refs. 5 and 13 for the effective sound speed is valid. There is generally an altitude, z_I, at which the effective sound speed is maximum. In the asymptotically constant approximation the effective sound speed profile is replaced with one that is constant, equal to the maximum sound speed, for altitudes z greater than z_I.

Let r be the horizontal range from a point source of unit amplitude. Let the source altitude be z_S and the receiver altitude be z. The modal approximation to the acoustic pressure amplitude at the receiver at angular frequency ω is given by

$$\hat{p}(r,z,\omega) = -\sqrt{\frac{i}{8\pi r}}\sum_{j=0}^{N(\omega)}\frac{1}{\sqrt{k_j(\omega)}}\psi_j(\omega,z_S)\psi_j(\omega,z)e^{ik_j(\omega)r}, \tag{1}$$

where $N(\omega)$ is the number of modes, $\psi_j(\omega,z)$ is the jth mode function, and $k_j(\omega)$ is the corresponding modal wave number. The modes satisfy

$$\left(\frac{d^2}{dz^2} + \left(\frac{\omega}{c_{\text{eff}}(z)} + i\alpha_{\text{atm}}\right)^2 - k_j^2\right)\psi_j = 0, \tag{2}$$

$$\int_0^\infty \psi_j(\omega,z)^2 dz = 1, \tag{3}$$

and

$$\left.\frac{d\psi_j(\omega,z)}{dz}\right|_{z=0} = -\frac{i\omega\rho_0}{\mathcal{Z}(\omega)}\psi_j(\omega,0), \tag{4}$$

where α_{atm} models the atmospheric attenuation,[9,10,19] ρ_0 is the mean density of the air, and $\mathcal{Z}$ is the ground impedance. Let ω_j be the cutoff frequency for the jth mode; for $\omega < \omega_j$ there are $j-1$ modes while for $\omega > \omega_j$ there are j modes.

The $j=0$ mode is identified with the ground-coupled surface mode.[5] The higher modes, those with $j>0$, will in this work be referred to as the refracted modes. The phase velocities c_j and attenuation coefficients α_j are related to the wave numbers k_j through

$$k_j(\omega) = \frac{\omega}{c_j(\omega)} + i\alpha_j(\omega). \tag{5}$$

The group velocity

$$c_{gj}(\omega) = \frac{1}{\text{Re}\ \partial k_j/\partial\omega} = \frac{c_j(\omega)}{1 - \dfrac{\omega}{c_j(\omega)}\dfrac{\partial c_j}{\partial\omega}}$$

is interpreted as the horizontal speed with which a wave packet, centered at frequency ω, and composed only of the jth mode, propagates.[20,21] The phase velocity generally decreases with increasing frequency so that the group velocity is less than the phase velocity. We note that some of the numerical results about the mode functions and their wave numbers obtained in Refs. 5 and 13 and below agree qualitatively with the analytical results presented in Refs. 6–8 and 16.

An impulsive signal may be obtained from Eq. (1) by inverse Fourier transform. Let $q(t)$ be the temporal waveform of an impulsive point source with spectrum

$$Q(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} q(t)e^{i\omega t}dt.$$

The pressure signal received at range r and altitude z is given by

$$p(r,z,t) = \int_{-\infty}^{\infty} Q(\omega)\hat{p}(r,z,\omega)e^{-i\omega t}d\omega. \tag{6}$$

Substituting Eq. (1) the pressure signal may be represented as a superposition of independently propagating dispersive modes

$$p(r,z,t) = -2\ \text{Re}\sqrt{\frac{i}{2\pi r}}\sum_{j=0}^{\infty}\int_{\omega_j}^{\infty}\frac{Q(\omega)}{\sqrt{k_j(\omega)}}\psi_j(\omega,z_S)\psi_j(\omega,z) \times e^{ik_j(\omega)r - i\omega t}d\omega. \tag{7}$$

Note that, as a function of frequency, the narrowband pressure amplitude Eq. (1) may be viewed as the transfer function for propagation through a horizontal distance r from a point at altitude z_S to a point at altitude z. For $|\omega| > |\omega_j|$ a transfer function for the jth mode may be defined by

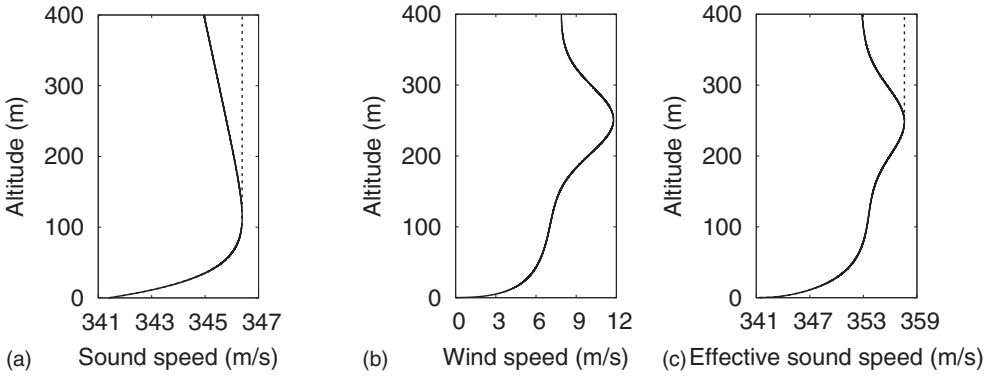


FIG. 2. The meteorological model. (a) The thermodynamic sound speed $c(z)$ as given by Eqs. (10) and (9). To be referred to as the "no-wind" model. (b) The model magnitude of the mean wind speed $v(z)$ as given by Eq. (11). (c) The down-wind effective sound speed profile. To be referred to as the "down-wind" model. The dotted lines in (a) and (c) indicate the asymptotically constant approximation.

$$\hat{p}_j(r,z,\omega) = -\sqrt{\frac{i}{8\pi r k_j(\omega)}}\psi_j(\omega,z_S)\psi_j(\omega,z)e^{ik_j(\omega)r}. \quad (8)$$

III. THE MODEL ENVIRONMENT

In this section some analytical forms are introduced to model the temperature and wind profiles and the ground impedance. For both the meteorology and the ground properties, the models used are intended to be simple, but sufficient to capture the gross features expected to be found in the field.

A. The meteorological model

On clear nights over land there is generally a temperature inversion near the ground: The temperature increases with altitude up to the top of the inversion. Above the inversion the temperature decreases, eventually decreasing linearly at a rate referred to as the lapse rate. The temperature gradient is positive at the ground, decreasing with increasing altitude, passing through 0 at the top of the inversion, and finally approaching the negative of the lapse rate at higher altitudes.

In this paper a simple exponential model for the temperature,

$$T(z) = T_G + \Delta(1 - \exp(-\alpha z)) - \Gamma z, \quad (9)$$

will be used.[1] Here T_G is the air temperature at the ground, Δ and α are parameters controlling the depth of the inversion, and Γ is taken to be the adiabatic lapse rate given approximately by 10 K/km.[1,2] The thermodynamic sound speed is given by

$$c(z) = \sqrt{\gamma R T(z)}, \quad (10)$$

where, for air, $\gamma = 1.4$ and $R = 287\ \mathrm{m^2/s^2\ K}$. In Fig. 2(a) the resulting model thermodynamic sound speed profile is plotted for the values $T_G = 290$ K, $\Delta = 10$ K, and $\alpha = 0.03/\mathrm{m}$.

Over perfectly flat ground the wind speed is expected to have the Bussinger–Dyer logarithmic plus linear form near the ground.[1,2] Let

$$v_{\mathrm{BD}}(z) = \begin{cases} 0 & \text{if } z < z_{\mathrm{stag}} \\ \ln\dfrac{z}{z_{\mathrm{stag}}} + \beta(z - z_{\mathrm{stag}}) & \text{if } z > z_{\mathrm{stag}} \end{cases}$$

be the Bussinger–Dyer model. The stagnation height z_{stag} is taken here to be 0.2 m. A form for β is given in literature, which depends on the Monin–Obukhov scale;[1,2] here the choice $\beta = 1$ will be made.

Above the temperature inversion there is the nocturnal jet. A Gaussian profile

$$v_J(z) = e^{-((z-z)_J/w_J)^2}$$

for the nocturnal jet is used. The jet altitude z_J is taken to be 250 m and the jet width w_J to be 66.7 m.

A model that interpolates between the low and high altitude forms is given by

$$v(z) = v_{\mathrm{ref}}\frac{\epsilon_{\mathrm{BD}} + v_{\mathrm{BD}}(z_{\mathrm{ref}})}{v_{\mathrm{BD}}(z_{\mathrm{ref}})}\left(\frac{v_{\mathrm{BD}}(z)}{\epsilon_{\mathrm{BD}} + v_{\mathrm{BD}}(z)} + a v_J(z)\right), \quad (11)$$

where, in this paper, the reference altitude z_{ref} is taken to be 2 m and the relative jet magnitude a to be 0.5. The parameter ϵ_{BD} controls the transition between the Bussinger–Dyer form and the nocturnal jet. It has been taken to be 10 m/s. In Fig. 2(b) the resulting model wind speed magnitude profile with $v_{\mathrm{ref}} = 2\ \mathrm{m/s}$ is plotted.

Simplifying further by assuming the wind direction to be horizontal and independent of altitude, the resulting effective sound speed is given by

$$c_{\mathrm{eff}}(z) = c(z) + \cos\theta\, v(z), \quad (12)$$

where θ is the angle the propagation direction makes with the wind direction. In Fig. 2(c) the resulting down-wind ($\theta = 0$) effective sound speed is plotted.

B. The ground impedance model

In this paper the effect of the ground on sound propagation is modeled through the impedance condition, as given in Eq. (4). The ground is assumed to be locally reacting in that the ground impedance depends only on frequency and not on angle of incidence. For frequencies greater than about 200 Hz this is known to be a good approximation.[22] A class of

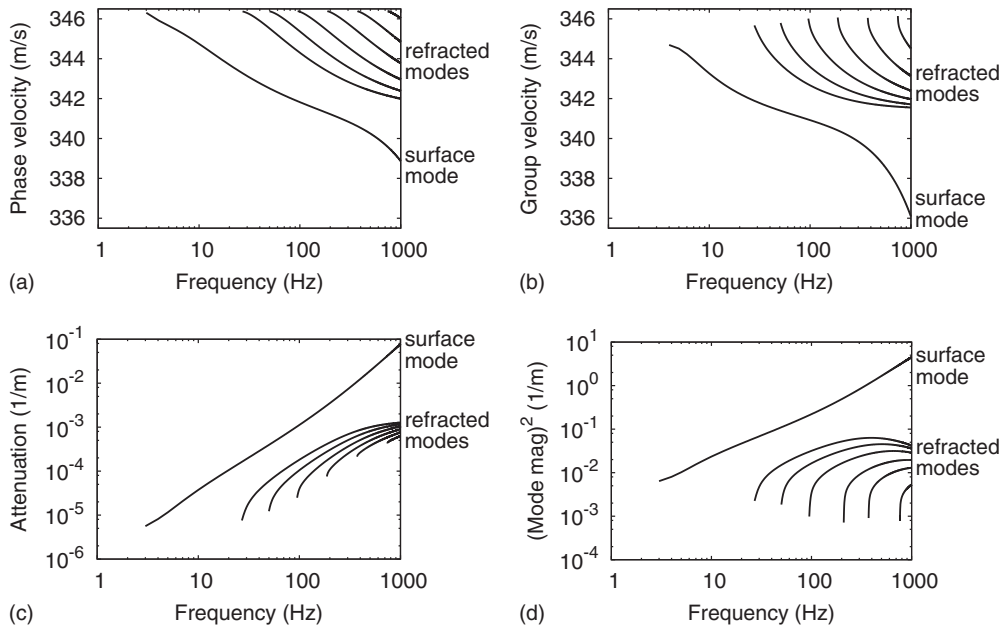


FIG. 3. The dispersion relations for the model without wind. The surface mode and refracted modes 1, 2, 4, 8, 16, and 32 are shown. (a) The phase velocity c_j. (b) The group velocity c_{gj}. (c) The attenuation coefficients α_j. (d) The magnitudes of the modes on the ground $|\psi_j(0)|$.

theoretical models, both fundamental and empirical, have been developed, which can be fit to the measured impedance over a broad frequency range,[17,22–24] and a standard method has been developed for measuring ground impedance in the field.[25] Note that the ground impedance is complex valued. The imaginary part represents the acoustic compliance of the ground and the real part represents the acoustic attenuation by the ground.

For frequencies lower than about 200 Hz, less is understood about the effect of the ground.[17,22] For sufficiently low frequencies, the locally reacting approximation for the ground impedance breaks down.[26] Measurements of ground impedance, assuming the ground to be locally reacting, have been reported down to about 20 (Ref. 27) and 50 Hz.[28] In all cases the imaginary part of the ground impedance is found to be much larger than the real part. It should be noted that theoretical models for ground in which the porosity decreases rapidly with increasing depth predict impedances with relatively large imaginary parts.[29]

No attempt will be made here to realistically model the ground. The ground will be assumed to be locally reacting. The ground impedance model used is very simple, yet sufficient to qualitatively reproduce the observations. The ground impedance is assumed to be

$$\frac{\mathcal{Z}(\omega)}{\rho_0 c(0)} = \zeta_0 e^{i\theta_0}\sqrt{\frac{f_0}{f}}, \tag{13}$$

where f is cyclic frequency, $c(0)$ is the sound speed on the ground, ζ_0 is the magnitude of the normalized ground impedance at a reference frequency f_0, and θ_0 is the phase angle between the real and imaginary parts. The $1/\sqrt{f}$ frequency dependence is typical of porous media. For the low frequencies of greatest significance for the pulse tail, the imaginary part of the ground impedance is expected to be considerably larger than the real part.[27–29] The phase angle, θ_0, is accordingly chosen close to $\pi/2$. In this paper the values $\zeta_0=25$ and $\theta_0=0.42\pi$ are used.

IV. PROPAGATION IN THE NO-WIND MODEL

In this section impulse propagation in the no-wind model is studied. Strictly speaking, the no-wind model corresponds to crosswind propagation, $\theta=\pi/2$ in Eq. (12), in the model presented in Sec. III A. In reality, the wind direction is rarely independent of altitude so that the notion of crosswind propagation is, to some degree, arbitrary.

A. The modal dispersion

To use Eq. (7) to obtain the propagated pressure signal the modal wave numbers $k_j(\omega)$, or equivalently the phase velocities $c_j(\omega)$ and attenuation coefficients $\alpha_j(\omega)$, and mode functions $\psi_j(\omega,z)$ are required. The wave numbers and mode functions are obtained numerically by discretizing the z axis and replacing the second derivative in Eq. (2) and the boundary condition (4) with a difference operator.[30] The resulting matrix is diagonalized in two steps. First the real part of the impedance is set to zero and Sturm theory[31] is used to find the eigenvalues of the resulting self-adjoint tridiagonal matrix. The real part of the impedance is then given its actual value and the technique of inverse iteration is used to find the desired eigenvalues and eigenvectors.[31,32]

The modal phase velocities c_j, group velocities c_{gj}, attenuation coefficients α_j, and mode magnitudes on the ground $|\psi_j(0)|$ are plotted over the frequency range 1–1000 Hz for the no-wind model in Fig. 3. In the asymptotically constant approximation the total number of modes at 1000 Hz is 44 in the no-wind model. The magnitude of the transfer function for ground to ground propagation at a range of 3 km is shown in Fig. 4. In Fig. 4(a) the transmission loss $TL=-20\log(\hat{p}(r,z,\omega))$ is plotted as a function of frequency at a

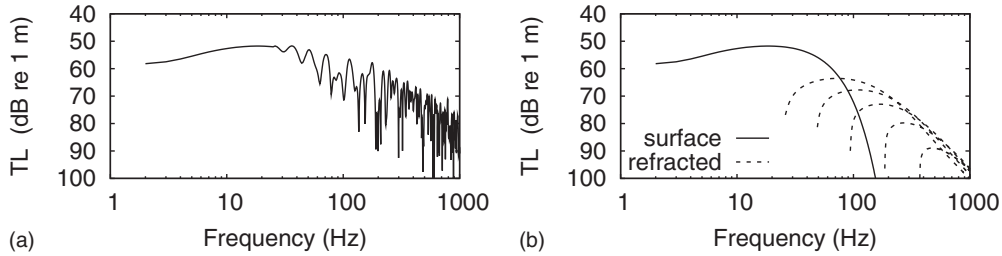


FIG. 4. The transmission loss as a function of frequency for ground to ground propagation at a fixed range of 3 km for the no-wind model. (a) The full transmission loss. (b) The transmission loss for the individual modes. The surface mode and every 2^{nth} refracted mode are shown.

fixed range $r=3$ km and fixed altitude $z=0$ m. In Fig. 4(b) the modal transmission losses $-20\log(\hat{p}_j(r,z,\omega))$ are plotted, again for $r=3$ km and $z=0$ m. Rather than showing all of the modes in Fig. 3 and in Fig. 4(b) only the surface mode and every 2^{nth} refracted mode; modes 1, 2, 4, 16, and 32; are displayed.

In Fig. 3 the surface mode stands out from the refracted modes as a special case.[5] The wave numbers of the refracted modes cluster together, getting closer with increasing mode number and frequency. The surface mode wave number remains separated from the refracted mode wave numbers, the separation growing with increasing frequency in both the real and imaginary parts. As a consequence, at a given frequency, the surface mode is more slowly propagating and much more highly attenuated, about an order of magnitude more so, than are the refracted modes.

In Fig. 4(a) the full transmission loss is seen to be intermediate between cylindrical spreading (35 dB at 3 km) and spherical spreading (70 dB at 3 km) across the frequency band of interest. As seen in Fig. 4(b), however, the transmission losses for the individual modes behave quite differently from the full transmission loss and the propagation of the surface mode component is qualitatively different from that of the refracted modes.

The surface mode has essentially no cutoff frequency but its transfer function rolls off rapidly with increasing frequency due to the severe attenuation of the surface mode by the ground. In this way the surface mode component may be thought of as a low pass filter, with its frequency band limited from above by the ground attenuation.

The refracted modes all have cutoff frequencies. The cutoff frequencies increase with increasing mode number. The high frequency behaviors of the refracted mode transmission losses are all similar to each other. This is because the refracted modes are attenuated primarily by the atmosphere, independent of the mode number. Thus, at high frequencies, they all roll off in roughly the same way, regardless of the mode number. The roll off is gentler than for the surface mode and begins at a much higher frequency. In this way the refracted mode components may be thought of as bandpass filters, with the frequency band for a mode limited from below by the cutoff frequency and from above by atmospheric attenuation. Since the attenuation becomes more and more severe with increasing range, while the cutoff frequency remains fixed, the spectrum of a single refracted mode becomes more and more narrowband as it propagates.

The following qualitative picture for broadband modal propagation emerges from the above discussion. The individual modal components cannot remain temporally localized as they propagate, their duration at any given range being given roughly by the inverse of their bandwidths. The computed group velocities in Fig. 3(b) show that the individual modes cannot be expected to separate from each other as they propagate. The sole exception is the surface mode. There is a frequency, about 40 Hz in the example considered in Fig. 3, above which the group velocity of the surface mode is less than that of any other modes. Above this frequency the surface mode arrives last and by itself.

It was argued in Ref. 5 that the behavior of the modal dispersion described above, and consequently of the modal transfer functions, is typical of propagation on clear and dry nights over a locally reacting ground. The separation of the surface mode wave numbers from those of the refracted modes is a consequence of the steep sound speed gradient near the ground. The clustering of the refracted mode wave numbers is a consequence of the shallow sound speed gradient up near the inversion. Finally, since the surface mode is generally much larger on the ground than are the refracted modes the influence of the ground on the surface mode is much greater than its influence on the refracted modes.

B. Propagation of an impulsive signal

To investigate the propagation of an impulse we consider the model impulse shown in Fig. 5(a). This impulse is obtained from by fitting the Fourier transform of the pulse shown in Fig. 1(a) for $\omega>0$ to

$$Q(\omega)=A\omega e^{-1/2(B\omega+(B\omega)^2)}. \tag{14}$$

The fitted function was multiplied by 20 to account for the distance to the source, divided by 2 to account for the approximate pressure doubling near the ground, and multiplied

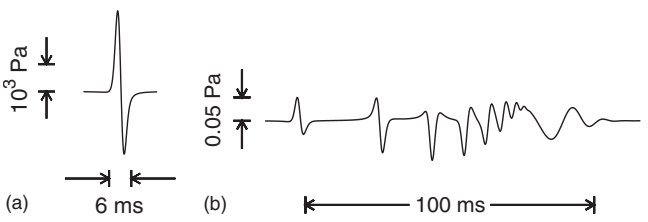


FIG. 5. An impulsive signal propagated 3 km from ground to ground in the no-wind model. (a) The waveform whose Fourier transform is Eq. (14) with $A=e^{-0.4\pi i}(1.153\times10^{-2}$ Pa) and $B=7.579\times10^{-4}$ s. (b) The waveform of (a) propagated to 3 km from the source.

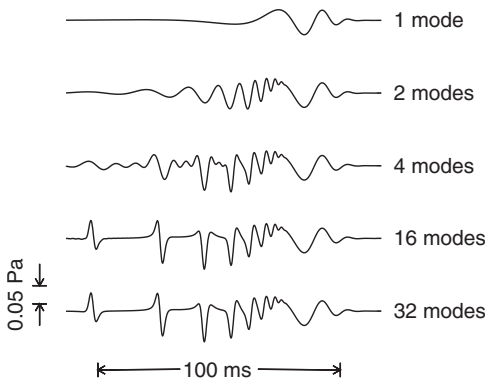


FIG. 6. The waveforms obtained at 3 km by restricting the modal sum (7) to the first 1, 2, 4, 16, and 32 modes.

by 4π to convert to a point source strength. One obtains $A = e^{-0.4\pi i}(1.153 \times 10^{-2}$ Pa) and $B = 7.579 \times 10^{-4}$ s. Performing an inverse Fourier transform on Q yields the waveform shown in Fig. 5(a). Substituting Q into Eq. (7) with $z_S = z = 0$ m and $r = 3000$ m yields the waveform shown in Fig. 5(b). Note the structure of the received signal: There are several distinct arrivals followed by a sequence of arrivals riding on top of, and leading into, a low frequency tail.

Since the surface mode propagates more slowly and attenuates high frequencies much more rapidly than do the refracted modes, one expects the low frequency tail to be the surface mode component of the pulse. It was noted in Sec. IV A that, as they propagate to long ranges, the individual refracted mode components never separate from each other temporally. As a single refracted mode propagates its bandwidth decreases and so that it is temporally broadened. It follows that the individual distinct arrivals at the beginning of the wave train cannot be expected to be individual refracted mode components but must be superpositions of many modes.

Figure 6 shows the waveforms obtained by restricting the modal sum (7) to 1 mode, the surface mode alone, and to 2, 4, 16, and 32 modes, and the surface mode plus the first 1, 3, 15, and 31 refracted modes, respectively. Indeed, the low frequency tail is precisely the surface mode component while the distinct initial arrivals require the superposition of many modes to be represented with any accuracy. Furthermore, the sequence of arrivals riding up on the leading edge of the surface mode component is seen to be the modulation resulting from the superposition of the first few refracted modes with the surface mode.

C. The ray paths

To understand the nature of the distinct initial arrivals we consider the geometrical acoustics approximation.[11,33] The ray that leaves the source position at launch angle ϕ_0 from horizontal will be parametrized by giving its altitude as a function of horizontal range r from the source,

$$z = \xi(\phi_0, r).$$

Since ground to ground propagation is being studied, only rays beginning on the ground will be considered. Thus $\xi(\phi_0, 0) = 0$. The ray paths are obtained numerically by solving the eikonal equation. Details are available in literature, for example, in Refs. 11, 17, 30, and 33.

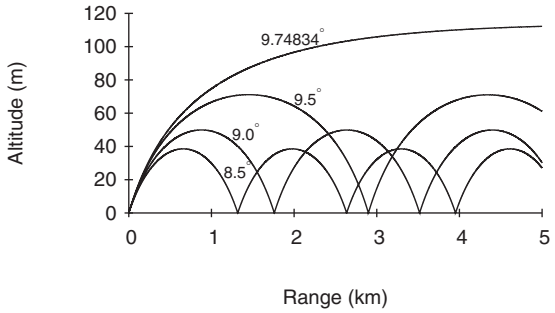


FIG. 7. The ray paths in the no-wind model with launch angles 8.5°, 9°, and 9.5° and with the critical launch angle $\phi_C \approx 9.748\,34°$.

Recall the altitude z_I, introduced in Sec. II, at which $c(z)$ has its maximum. There is a unique critical launch angle, to be denoted here by ϕ_C, for which the corresponding ray altitude $\xi(\phi_C, r)$ is asymptotic to z_I as follows:

$$\lim_{r \to \infty} \xi(\phi_C, r) = z_I. \tag{15}$$

In the model shown in Fig. 2(a) one has $\phi_C \approx 9.75°$.

Launch angles greater than ϕ_C correspond to rays that go up into the atmosphere and never return to earth. Launch angles less than ϕ_C correspond to rays that are refracted back to earth by the sound speed gradient. On striking the ground, the ray is reflected back up into the atmosphere, is then refracted back toward the ground, and so on. Several rays in the no-wind model with launch angles less than and equal to ϕ_C are depicted in Fig. 7.

It was hypothesized in Ref. 6 that, for ground to ground propagation, the arrivals can be classified by the number of ground reflections encountered along their ray paths. The first arrival corresponds to a ray path that does not touch the ground before reaching the receiver. The second arrival corresponds to a ray path that touches the ground once and so on. This hypothesis is based on the behavior of the rays for ground to ground propagation as shown in Fig. 8. Launch angles for the ray paths with no, one, two, and three ground reflections are shown in Fig. 8(a) as a function of horizontal range, r, and in Fig. 8(b) as a function of travel time, τ. Figure 8(c) shows the celerity, r/τ, plotted as a function of range r. At a fixed range, the celerity decreases with increasing number of ground reflections so that the greater the number of ground reflections, the greater the predicted travel time. Note that for a fixed number of ground reflections, both range and travel time are monotone increasing as functions of the launch angle.

In Fig. 9 the waveform obtained from modal expansion is compared to the arrival structure predicted by ray theory. The horizontal axis shows time $t - r/c(z_I)$ relative to propagation at the maximum sound speed; $c(z_I)$ with $t = 0$ is chosen to be the zero-crossing time for the source waveform [Fig. 5(a)]. The solid curves are the results of propagation to ranges r equal to 2, 3, 4, 5, and 6 km. For the sake of visibility in the figure spherical spreading has been removed: signal amplitudes have been multiplied by range. The dashed

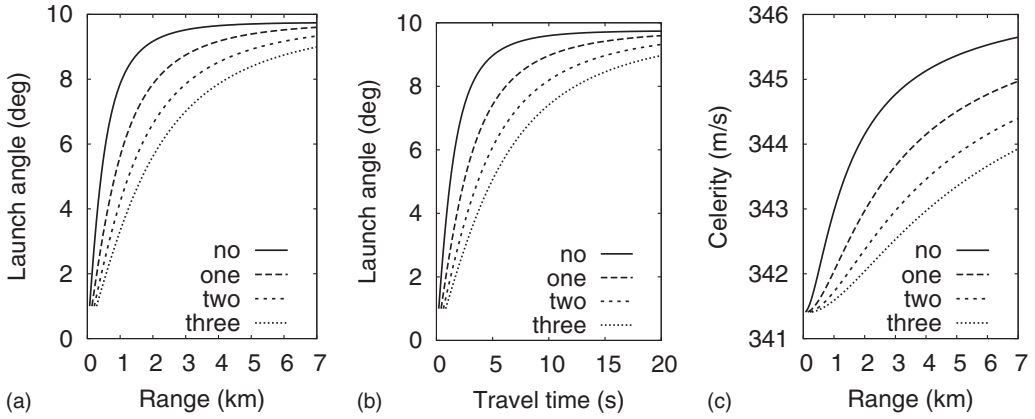


FIG. 8. Properties of ray paths for ground to ground propagation in the no-wind model for no, one, two, and three ground reflections. (a) Launch angle as a function of range. (b) Launch angle as a function of travel time. (c) Celerity (range divided by travel time) as a function of range.

curves are the ranges as functions of arrival time as predicted by ray theory for the no, one, two, three, and four times ground reflected arrivals.

The correspondence between the time of the first few isolated arrivals at the beginning of the wave train and the arrival times predicted by ray theory is quite good.[34] The first arrival corresponds to the ray with no ground reflections, the second to the ray with one ground reflection, and so on until arrivals begin to overlap as the tail forms and ray theory becomes invalid. Toward the end of the waveform, where the signal is the superposition of a small number of modes, the ray-theoretic travel times are not meaningful.

It is shown in the Appendix that a simple caustic is necessarily encountered between each ground bounce and the subsequent turning point. The significance of this is that as a ray passes through a caustic the amplitude of the corresponding acoustic signal undergoes a phase shift of 90°.[17,33] Recall that shifting the phase of a time series by 90° is equivalent to taking the Hilbert transform of the signal. Ray theory predicts that the nth arrival in the pulse train corresponds to a ray path, which has been reflected off the ground $n-1$ times and passed through $n-1$ simple caustics.

In the geometrical acoustics approximation, the effect of the ground reflection on the pulse can be approximated by convolving the waveform with the spherical wave reflection coefficient,[17,35,36] while the effect of passing through the caustic is accounted for by taking the Hilbert transform of the waveform. In Fig. 10 the source waveform [in Fig. 10(a)], the source waveform filtered once through the Hilbert transform and the spherical wave reflection coefficient [in Fig. 10(b)], and the source waveform filtered twice through the Hilbert transform and the spherical wave reflection coefficient [in Fig. 10(c)] are displayed. Comparing the results in Fig. 10 to the propagated pulse in Fig. 5(b) shows the predictions of geometric acoustics to be qualitatively true for the first few arrivals of the received pulse train. It should be noted that if the waveform is Hilbert transformed, but not filtered through the spherical wave reflection coefficient, the resulting waveforms do not resemble the arrivals of the propagated impulse.

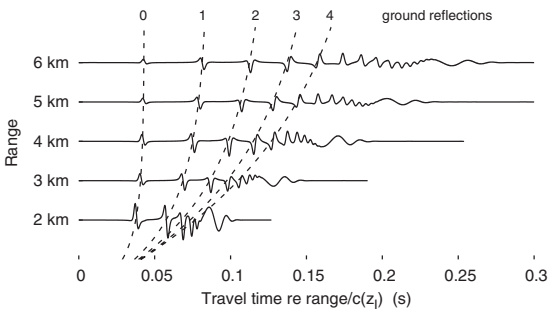


FIG. 9. Propagation with no-wind of the impulse shown in Fig. 5(a) in steps of 1 km from 2 to 6 km. The horizontal axis shows $t-r/c(z_l)$ where r is range and $t=0$ is the zero-crossing time for the source waveform.

D. The influence of the ground impedance

We end this section by considering how different ground impedances affect the waveform of the propagated signal. Physically, one expects that the more contact a component of the received pulse has had with the ground the greater the influence of the ground on that component. One expects that the first arrival, which suffers no ground reflections, is insensitive to the ground, that the more ground reflections a ray experiences, the greater the effect, and that the effect on the surface mode tail should be significant.

Indeed, under a small change $\delta\mathcal{Z}$ in the ground impedance, using Eq. (19) of Ref. 13, approximating $k_j \approx \omega/c(0)$ and introducing the normalized impedance $\hat{\mathcal{Z}}=\mathcal{Z}/\rho_0 c(0)$, the variation of the wave number of the jth mode is given by

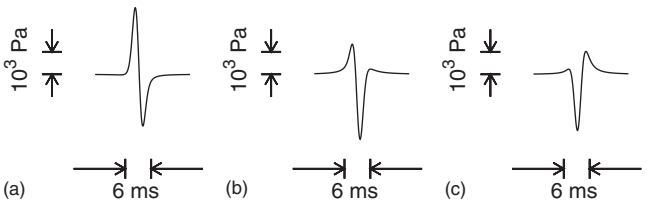


FIG. 10. (a) The point source waveform. (b) The Hilbert transform of the ground reflected point source waveform. (c) Two Hilbert transforms of the twice reflected point source waveform.

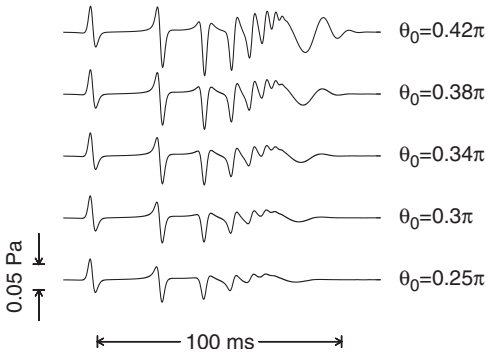


FIG. 11. The impulse propagated to 3 km in the no-wind model for ground impedances of the form (13) for a variety of complex phases θ_0.

$$\frac{\delta k_j}{\delta \hat{\mathcal{Z}}} \approx \frac{i}{2\hat{\mathcal{Z}}^2}\psi_j(\omega,0)^2. \tag{16}$$

In particular, the variation of δk_j is proportional to the square of the jth mode magnitude on the ground. Since the magnitude of the square of the surface mode is an order of magnitude larger on the ground than that of the refracted modes, see Fig. 3(d), a change in ground properties has a much greater effect on the surface mode components than it does on the refracted mode components.

In Fig. 11 the effect of changing the ground impedance model in Eq. (13) is studied. The magnitude of $\mathcal{Z}$ is kept fixed while the phase is varied from 0.42π (about 74.6°) to 0.25π (45°). As the phase of the ground impedance decreases the ground becomes less compliant and more resistive. The surface mode tail is changed dramatically by the change in impedance. As the phase of the impedance decreases, the surface mode component decreases in amplitude and arrives earlier. The effect on the initial arrivals is less. Their arrival times do not change significantly; however, the amplitudes of the initial arrivals decrease as the ground becomes more resistive, the more so the greater the number of ground reflections.

V. DOWN-WIND PROPAGATION AT NIGHT

The down-wind model differs from the no-wind model in two ways. First, the duct in the temperature inversion is deepened by the low-altitude winds. Second, the nocturnal jet provides ducting at higher altitudes. The effective sound speed gradient is steepest near the ground so that the pulse tail will be similar to that of the no-wind case, only of longer duration. At long ranges the structure of the initial arrivals is, however, more complex than in the no-wind model due to refraction by the nocturnal jet.

In this work only the simple single-jet wind model shown in Fig. 2(b) will be considered. In this model the effective sound speed profile has, as in the no-wind model, a unique local maximum at an altitude z_l. If the nocturnal jet is somewhat higher or weaker, or if there is a low altitude jet, then the effective sound speed profile can have two maxima. This case will not be treated explicitly here except to say that the behavior is similar to the case considered. Fine structure in the wind speed profile also will not be considered.

A. The modal dispersion and the pulse tail

The modal wave numbers and mode functions are determined in the down-wind model in the same way as in the no-wind case. The modal phase velocities c_j, group velocities c_{gj}, attenuation coefficients α_j, and mode magnitudes on the ground $|\psi_j(0)|$ are plotted over the frequency range 1–1000 Hz for the down-wind model in Fig. 12. In the asymptotically constant approximation the total number of modes at

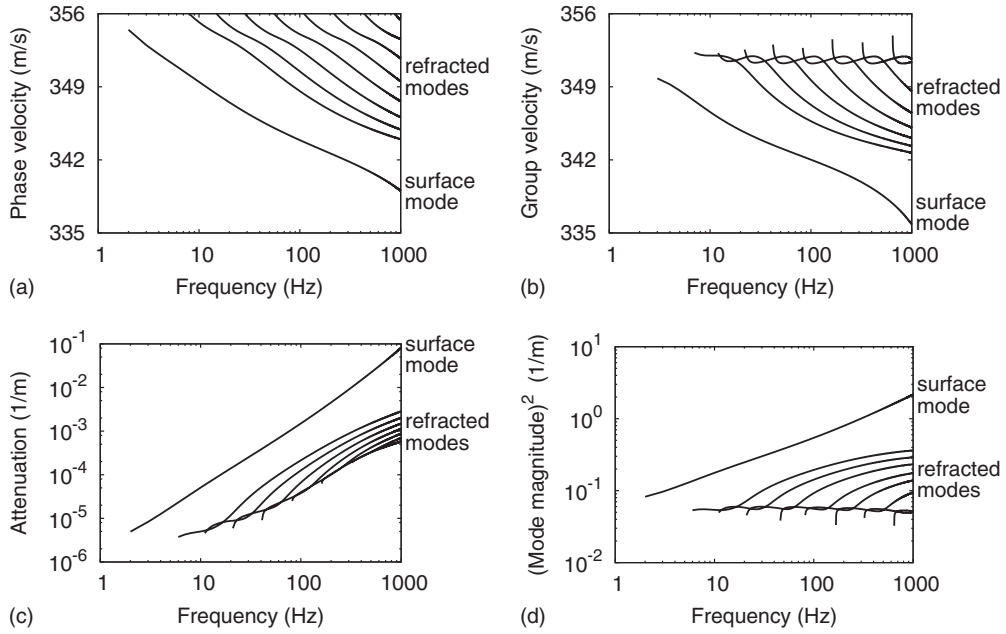


FIG. 12. The dispersion relations for the down-wind model. The surface mode and refracted modes 1, 2, 4, 8, 16, 32, 64, and 128 are shown. (a) The phase velocity c_j. (b) The group velocity c_{gj}. (c) The attenuation coefficients α_j. (d) The mode magnitude $|\psi_j(0)|$.

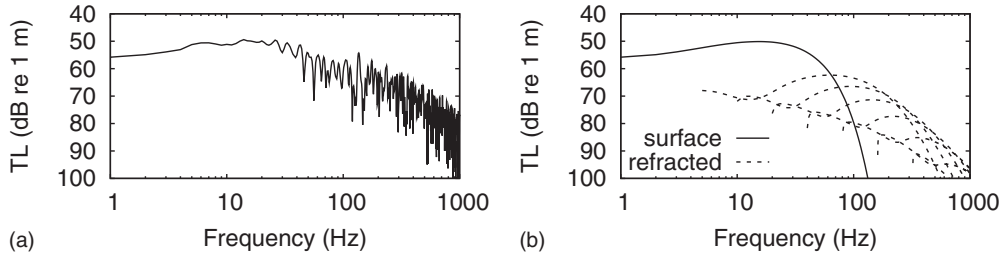


FIG. 13. The ground to ground transfer function at 3 km for the downwind model. (a) The full transfer function. (b) The transfer functions restricted to individual modes: the surface mode and every 2^{nth} refracted mode.

1000 Hz is 199 in the down-wind model (versus 44 in the no-wind model). The surface mode and the refracted modes 1, 2, 4, 16, 32, 64, and 128 are displayed.

As in the no-wind case, the wave numbers of the refracted modes cluster together, getting closer with increasing mode number and frequency, while the surface mode wave number remains separated from the refracted mode wave numbers, the separation growing with increasing frequency. Again, the surface mode is significantly more slowly propagating and much more highly attenuated than the refracted modes. The dispersion curves are not as smooth as in the no-wind case, however. They show a transition from ducting by the nocturnal jet near a mode's cutoff frequency to ducting by the temperature lapse at higher frequencies.

The magnitude of the transfer function for ground to ground propagation at a range of 3 km is shown in Fig. 13. In Fig. 13(a) the transmission loss $-20\log(\hat{p}(r,3,\omega))$ is plotted as a function of frequency for $r=3$ km and $z=0$ m. As in the no-wind case, the transmission is intermediate between cylindrical spreading (35 dB) and spherical spreading (70 dB) across most of the frequency band. In Fig. 13(b) the modal transmission losses $-20\log(\hat{p}_j(r,z,\omega))$ are plotted, again for $r=3$ km and $z=0$ m. As in the no-wind case, the propagation of the surface mode is qualitatively different from that of the refracted modes. In addition to propagating more slowly than the refracted modes, the surface mode is a low pass filter. It has no cutoff frequency and its frequency band is limited from above by ground attenuation. Each refracted mode is again a bandpass filter, with its frequency band limited from below by its cutoff frequency and from above by atmospheric attenuation. As opposed to the no-wind case, the transfer functions for the refracted modes in the down-wind model show two frequency bands, a lower band representing propagation in the duct formed by the nocturnal jet and a higher frequency band representing propagation in the temperature inversion.

In Fig. 14 the fourth refracted mode is considered in detail. In Fig. 14(a) the transmission loss for mode 4 is plotted as a function of frequency at ranges of 3 and 5 km. The two frequency bands are clearly visible. The lower band is centered at about 25 Hz, and the higher band at about 140 Hz at 3 km, and at about 120 Hz at 5 km. The mode functions for the fourth refracted mode at 25 and 140 Hz are plotted over the effective sound speed profile in Fig. 14(b). The ver-

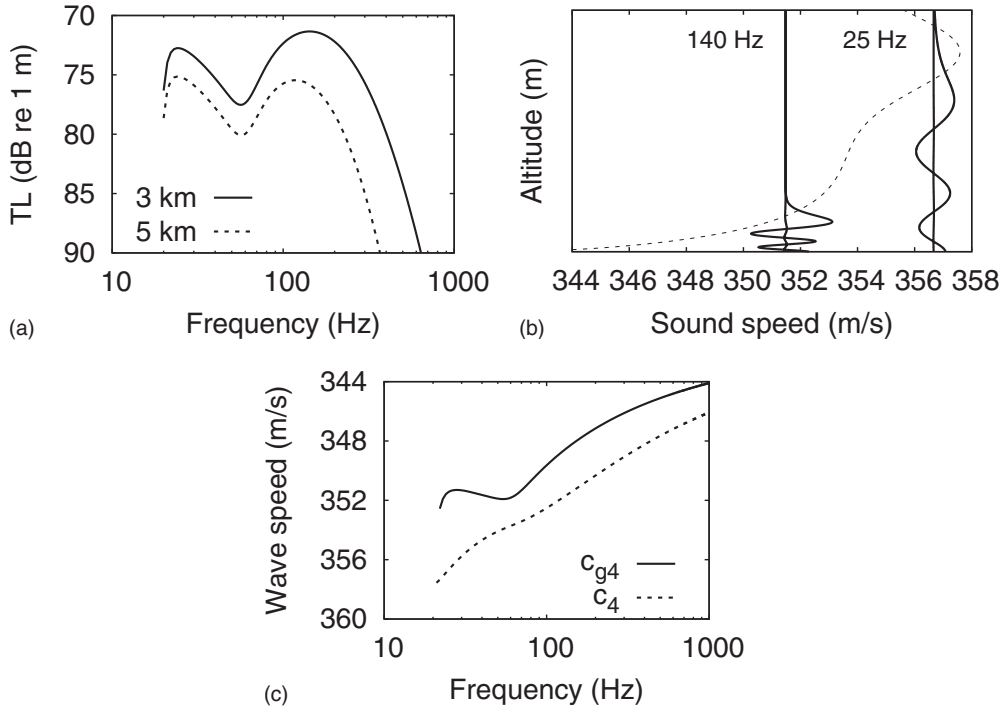


FIG. 14. The propagation of the fourth refracted mode. (a) The modal transmission loss at 3 and 5 km. (b) The mode functions at 25 and 140 Hz shifted by the phase velocity and plotted (solid lines) over the effective sound speed profile (dashed line). (c) The phase and group velocities.

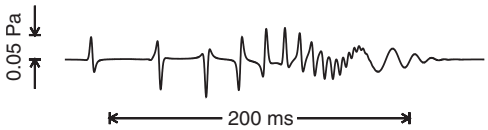


FIG. 15. The waveform of Fig. 5(a) propagated to 3 km from the source in the downwind model.

tical axes on which the two mode functions are plotted have been shifted to their respective phase velocities to show how the mode fits into the effective sound speed profile. Figure 14(b) demonstrates that the lower frequency band does indeed represent ducting by the nocturnal jet, while the higher frequency band represents ducting by the inversion. In Fig. 14(c) the phase and group velocities are plotted. In the downwind model the group velocities of the refracted modes are not monotonic as functions of frequency. In particular, at a fixed range, up to three distinct frequency components of a given refracted mode can arrive simultaneously.

Finally, the waveform, synthesized by modal expansion, of the signal from the source shown in Fig. 5(a) received at a range of 3 km is shown in Fig. 15. The waveform is similar to that obtained in the no-wind case except that its duration is twice as long. As in the no-wind case, at long ranges from the source, the waveform of an impulsive signal consists of several distinct arrivals, built up by the superposition of many modes, followed by a low frequency tail made up of fewer and fewer modes until, at the end of the pulse, only the surface mode is left.

B. The ray paths and the nocturnal convergence zones

As in the no-wind model, to understand the structure of the initial arrivals in the signal received at long ranges from an impulsive source, geometrical acoustics will be used. In the simple single-jet wind model shown in Fig. 2(b), the effective sound speed profile has, as in the no-wind model, a unique local maximum at an altitude z_I. Correspondingly, there is again a unique critical launch angle, ϕ_C, for which Eq. (15) is satisfied. For the model shown in Fig. 2(c), one has $\phi_C \approx 17.31417°$. In contrast to the no-wind model, in this downwind model, the range at which a ray strikes the ground

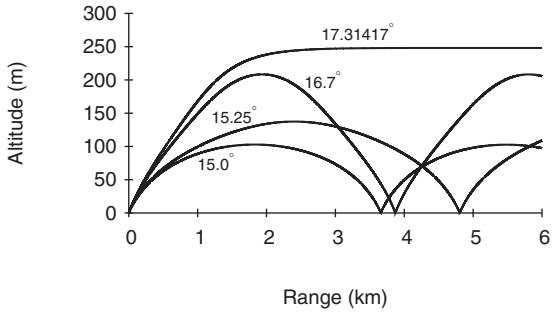


FIG. 16. The ray paths in the downwind model with launch angles 15°, 15.25°, and 16.7° and with the critical launch angle $\phi_C \approx 17.31417°$.

for the jth time is not strictly increasing with increasing launch angle. In Fig. 16 several rays with launch angles less than and equal to ϕ_C are shown. Note that the ray with a launch angle of 16.7° strikes the ground closer to the source than does the ray with a launch angle of 15.25°.

In Fig. 17 the behavior of the rays for ground to ground propagation in the downwind model is studied. As in Fig. 8, Figs. 17(a) and 17(b) show the launch angle plotted as a function of horizontal range, r, and travel time, τ, respectively, for no, one, two, and three ground reflections and Fig. 17(c) shows the celerity, r/τ, plotted as a function of range r.

In this case, the ray theory for ground to ground propagation is not as straightforward as it is in the no-wind model. Neither range nor travel time is a monotone function of the launch angle. As a consequence there are ranges at which arrivals refracted predominantly by the nocturnal jet and arrivals refracted predominantly by the temperature inversion coincide. The arrivals corresponding to ray paths with no ground reflections will be referred to as direct-refracted arrivals.[37] In the model shown in Fig. 2(c) there is a segment of ranges, from about 3.8 to 4.8 km, for which there are three distinct direct-refracted arrivals. For other ranges there is only one. Similarly, at $j+1$ times these ranges, one finds three distinct j-times ground reflected ground to ground ray paths.

The regions in which there are three j-times ground reflected arrivals will be referred to as the jth nocturnal convergence zone. Only the first convergence zone will be considered here. The ray-theoretic celerity for the direct-

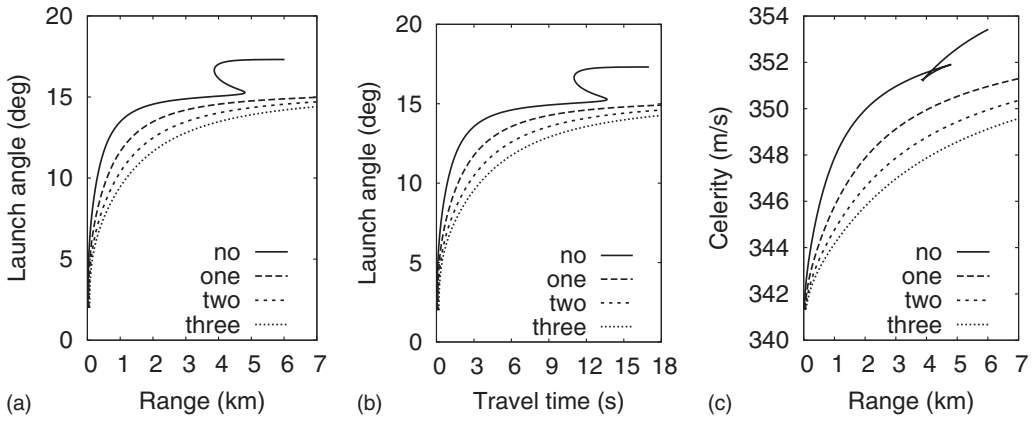


FIG. 17. Properties of ray paths for ground to ground propagation in the downwind model for no, one, two, and three ground reflections. (a) Launch angle as a function of range. (b) Launch angle as a function of travel time. (c) Celerity (range divided by travel time) as a function of range.

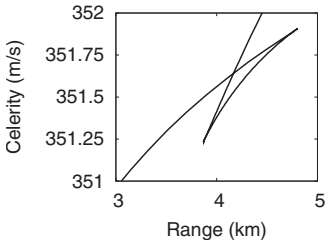


FIG. 18. Detail of the ray-theoretic celerity for the direct-refracted arrivals in the first convergence zone.

refracted ray paths in the first convergence zone is plotted in Fig. 18. It shows that the three arrivals are nearly coincident. Thus, in this region, there is a convergence of arrivals from the inversion with arrivals from the nocturnal jet.

Figure 19 shows the propagated waveforms, as synthesized by the modal expansion, compared to the arrival times predicted by ray theory. As in Fig. 9 the horizontal axis shows time relative to the shortest possible travel time in the model, $t-r/c_{\mathrm{eff}}(z_I)$. Time $t=0$ is again chosen to be the zero-crossing time for the source waveform in Fig. 5(a). The solid curves show the waveforms at a set of ranges r from 2.5 to 5.5 km, in steps of 0.5 km. Again, spherical spreading has been removed in that waveform amplitudes have been multiplied by range. The dashed curves are the ranges as functions of arrival time as predicted by ray theory for the no to five times ground reflected arrivals.

At shorter ranges, up to about 3 km or so, ground to ground propagation downwind is qualitatively similar to propagation in the no-wind model. The waveforms have longer duration, due to the duct being deeper, but the arrival structure is as in the no-wind case. At longer ranges the arrival structure changes significantly. In the convergence zone the first arrival from the nocturnal jet overtakes the first arrival from the inversion. The two arrivals converge, producing an anomalously large signal, and then separate.

Note that ray theory fails to correctly predict both the range at which the direct-refracted arrival from the nocturnal jet first appears, as well as the range to which the direct-refracted arrival from the temperature inversion persists. In Fig. 19 the direct-refracted arrival from the nocturnal jet is visible in the synthesized waveform at 3.5 km (at time 0.2 s). This arrival does not appear in the ray theory prediction (the dashed line) until about 3.9 km. Similarly, the direct-refracted arrival from the temperature inversion is still visible in the synthesized waveform at 5.5 km while it vanishes from the ray theory prediction by about 4.8 km.

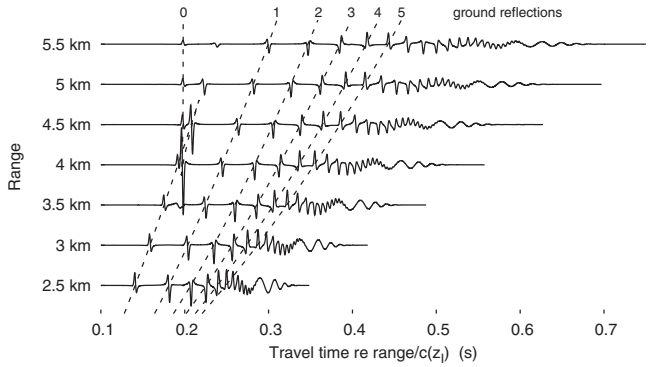


FIG. 19. Propagation downwind of the impulse shown in Fig. 5(a) in steps of 500 m from 2.5 to 5.5 km. The horizontal axis shows $t-r/c(z_I)$ where r is range and $t=0$ is the zero-crossing time for the source waveform.

It should be noted that the results presented here have been obtained for a particular meteorological model. The question arises of the generality of these results. Convergence zones will occur for down-wind propagation on clear nights whenever there are elevated wind jets. The convergence zones arise out of a competition between path length and sound speed. Signals ducted by an elevated jet will propagate faster than signals ducted by the temperature inversion but have longer propagation paths. The convergence zone is the region in which these two competing factors balance each other. Since nocturnal jets are always observed in fair weather over flat land,[3] such convergence zones should be common.

VI. CONCLUSIONS

A theoretical study of ground to ground propagation of impulsive sound outdoors under strongly ducted nocturnal conditions has been undertaken. A full wave solution is obtained using Fourier synthesis and a modal expansion. The influence of temperature, wind, and ground impedance on the arrival structure has been elucidated. At long ranges from the source the signal is seen to develop a characteristic structure: It generally begins with several distinct arrivals and ends with a low frequency narrowband tail.

The low frequency tail is the ground-coupled surface mode. It has been shown to be very sensitive to the ground surface over which it propagates. The initial arrivals have been identified by comparing the predictions of ray theory with those of the full wave solution. At moderate ranges, the jth arrival can be identified with the ray path that has undergone $j-1$ ground reflections. The waveforms of these initial arrivals are predicted by ray theory.

At longer ranges, the classification of the arrivals by the number of ground reflections breaks down. Elevated wind jets, in particular the nocturnal jet, can have a significant effect on the arrival structure and on the amplitudes of the arrivals. There are downwind convergence zones in which arrivals from the nocturnal jet and from the nocturnal temperature inversion converge and then separate. The convergence leads to anomalously large signals. In a convergence zone, ray theory predicts neither arrival structure nor waveform correctly.

ACKNOWLEDGMENTS

We gratefully acknowledge funding from the U. S. Army TACOM-ARDEC at Picatinny Arsenal, NJ. We would also like to thank Sergey Kulichkov, Igor Chunchuzov, and Richard Raspet for useful discussions. We thank Jin So for a critical reading of this manuscript.

APPENDIX: THE CAUSTICS IN THE NO-WIND MODEL

The analysis in Sec. IV C makes use of the fact that when a ray passes through a caustic the corresponding signal undergoes a 90° phase shift. In this appendix it is shown that,

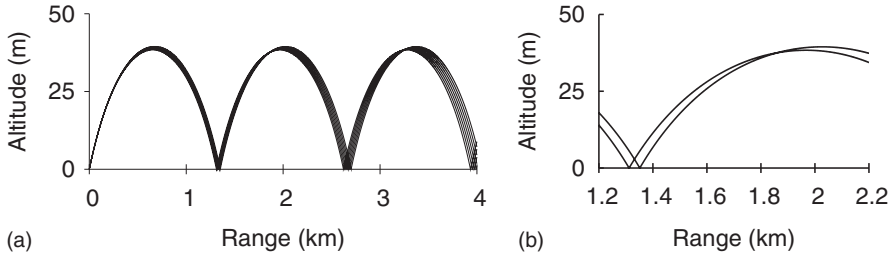


FIG. 20. (a) A cone of ray paths about 8.53° for the no-wind model. (b) A portion of the ray paths with launch angles 8.49° and 8.55°.

in the no-wind model, a caustic is necessarily encountered between each ground reflection and the subsequent turning point. The general theory for the location of caustics developed in Refs. 38 and 39 requires that the ray paths considered not intersect an interface. Hence, this theory cannot be applied here without modification.

In Fig. 20(a) a narrow cone of rays is shown with launch angles $\phi_0 \approx 8.53°$. Caustics, understood as points at which $\partial\xi/\partial\phi_0 = 0$,[38,39] are visible at 1.86 and 3.24 km. In Fig. 20(b) the first ground reflection and crossing point for the rays with launch angles 8.49° and 8.55° are shown. Consider ray paths $\xi(\phi_0, r)$ and $\xi(\phi_0 + \delta, r)$ with δ positive and small. Consider the jth ground reflection. The range at which $\xi(\phi_0, r)$ strikes the ground is less that that at which $\xi(\phi_0 + \delta, r)$ strikes the ground while the maximum altitude, the turning point, reached by $\xi(\phi_0, r)$ is less than that reached by $\xi(\phi_0 + \delta, r)$. Between the jth ground reflection and the $(j+1)$th turning point is a range $r(\delta)$ at which the rays cross as follows:

$$\xi(\phi_0 + \delta, r(\delta)) = \xi(\phi_0, r(\delta)). \tag{A1}$$

Note that the crossing point $r(\delta)$ lies between the point of ground reflection and the turning point.

It will be assumed that the limit crossing point $r(0)$ exists. Consider the partial derivative

$$\frac{\partial\xi(\phi_0, r(0))}{\partial\phi_0} = \lim_{\delta\to 0} \frac{\xi(\phi_0 + \delta, r(0)) - \xi(\phi_0, r(0))}{\delta}.$$

One finds, using Taylor expansion,

$$\begin{aligned}
&\xi(\phi_0 + \delta, r(0)) - \xi(\phi_0, r(0)) \\
&\quad = \xi(\phi_0 + \delta, r(\delta) - r'(0) \\
&\quad \times \delta + \ldots) - \xi(\phi_0, r(\delta) - r'(0) \\
&\quad \times \delta + \ldots) = \xi(\phi_0 + \delta, r(\delta)) - \xi(\phi_0, r(\delta)) + \mathcal{O}(\delta^2).
\end{aligned}$$

It follows from Eq. (A1) that

$$\frac{\partial\xi(\phi_0, r(0))}{\partial\phi_0} = 0$$

so that $r(0)$ is indeed a caustic for $\xi(\phi_0, r)$.

[1] R. B. Stull, *An Introduction to Boundary Layer Meteorology* (Kluwer, Dordrecht, 1991).
[2] J. R. Garratt, *The Atmospheric Boundary Layer* (Cambridge, Melbourne, 1992).
[3] C. Whiteman, X. Bian, and S. Zhong, "Low-level jet climatology from enhanced rawinsonde observations at a site in the southern great plains," J. Appl. Meteorol. **36**, 1363–1376 (1997).
[4] I. Chunchuzov, S. Kulichkov, A. Otrezov, and V. Perepelkin, "Acoustic pulse propagation through a fluctuating stably stratified atmospheric boundary layer," J. Acoust. Soc. Am. **117**, 1868–1879 (2005).
[5] R. Waxler, C. L. Talmadge, S. Dravida, and K. E. Gilbert, "The near-ground structure of the nocturnal sound field," J. Acoust. Soc. Am. **119**, 86–95 (2006).
[6] I. P. Chunchuzov, G. A. Bush, and S. N. Kulichkov, "On acoustical impulse propagation in a moving inhomogeneous atmospheric layer," J. Acoust. Soc. Am. **88**, 455–461 (1990).
[7] I. Chunchuzov and A. Otrezov, "On sound propagation in the atmospheric boundary layer," in Proceedings of the Eighth International Symposium on Acoustic Remote Sensing and Associated Techniques of the Atmosphere and Ocean, Moscow, 1996.
[8] I. Chunchuzov, A. Otrezov, and I. Petenko, "On inclined sounding of the stable stratified atmospheric boundary layer," in Proceedings of the Eighth International Symposium on Acoustic Remote Sensing and Associated Techniques of the Atmosphere and Ocean, Moscow, 1996.
[9] H. E. Bass, L. C. Sutherland, J. Piercy, and L. Evans, in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1984).
[10] H. E. Bass, L. C. Sutherland, and A. J. Zuckerwar, "Atmospheric absorption of sound: Update," J. Acoust. Soc. Am. **88**, 2019–2021 (1990).
[11] A. D. Pierce, *Acoustics* (Acoustical Society of America, Woodbury, NY, 1989).
[12] D. G. Albert, "Observations of acoustic surface wave in outdoor sound propagation," J. Acoust. Soc. Am. **113**, 2495–2500 (2003).
[13] R. Waxler, "A vertical eigenfunction expansion for the propagation of sound in a downward refracting atmosphere over a complex impedance plane," J. Acoust. Soc. Am. **112**, 2540–2552 (2002).
[14] D. C. Pridmore-Brown, "Sound propagation in a temperature- and wind-stratified medium," J. Acoust. Soc. Am. **34**, 438–443 (1962).
[15] O. A. Godin, "An effective quiescent medium for sound propagating through an inhomogeneous, moving fluid," J. Acoust. Soc. Am. **112**, 1269–1275 (2002).
[16] V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E & FN Spon, London, 1997).
[17] E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer, Dordrecht, 2001).
[18] R. Waxler, "Modal expansions for sound propagation in the nocturnal boundary layer," J. Acoust. Soc. Am. **115**, 1437–1448 (2004).
[19] ANSI, *Method of Calculation of the Absorption of Sound by the Atmosphere*, (American National Standards Institute, New York, 1995), Catalog No. ANSI S1.26-26-1995.
[20] J. W. S. Rayleigh, *Theory of Sound I and II* (Dover, New York, 1945).
[21] L. Brillouin, *Wave Propagation and Group Velocity* (Academic, New York, 1960).
[22] K. Attenborough, K. M. Li, and K. Horoshenkov, *Predicting Outdoor Sound* (Spon, London, 2001).
[23] K. Attenborough, "Acoustical impedance models for outdoor ground surfaces," J. Sound Vib. **99**, 521–544 (1985).
[24] K. Attenborough, "Ground parameter information for propagation modeling," J. Acoust. Soc. Am. **92**, 418–427 (1992).
[25] ANSI, *Template Method for Ground Impedance* (American National Standards Institute, New York, 2004), Catalog No. ANSI S1.18-1999 (R2004).
[26] C. Madshus, F. Lovholt, A. Kania, L. Hole, K. Attenborough, and S. Taherzahdeh, "Air-ground interaction in long range propagation of low frequency sound and vibration: Field tests and model verification," Appl. Acoust. **66**, 553–578 (2005).
[27] G. A. Daigle and M. R. Stinson, "Impedance of grass-covered ground at low frequencies measured using a phase difference technique," J. Acoust. Soc. Am. **81**, 62–68 (1987).
[28] A. Zuckerwar, "Acoustic ground impedance meter," J. Acoust. Soc. Am.

73, 2180–2186 (1983).
[29]R. Raspet and J. M. Sabatier, "The surface impedance of grounds with exponential porosity," J. Acoust. Soc. Am. **99**, 147–152 (1996).
[30]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, New York, 2000).
[31]J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 2nd ed. (Springer, New York, 1991).
[32]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, New York, NY, 1992).
[33]L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media II* (Springer, New York, 1999).
[34]An analogous result for narrowband signals was obtained in Ref. 35 for the case of a linear sound speed profile.
[35]R. Raspet, A. L'Espérance, and G. A. Daigle, "The effect of realistic ground impedance on the accuracy of ray tracing," J. Acoust. Soc. Am. **97**, 154–158 (1995).
[36]A. L'Espérance, P. Herzog, G. A. Daigle, and J. Nicolas, "Heuristic model for outdoor sound propagation based on an extension of the geometrical ray theory in the case of a linear sound speed profile," Appl. Acoust. **37**, 111–139 (1992).
[37]The term "direct-refracted" is used to distinguish these arrivals from the surface mode arrival, which might be considered a "direct-ground-coupled" arrival.
[38]W. A. Kinney and A. D. Pierce, "Caustics and the spreading of adjacent acoustic rays," J. Acoust. Soc. Am. **67**, 1145–1148 (1980).
[39]J. G. Watson, W. L. Siegmann, and M. J. Jacobson, "Refracted ray characteristics from the geometrical spreading equation for depth-dependent media," J. Acoust. Soc. Am. **70**, 583–590 (1981).

Near- to far-field characteristics of acoustic radiation through plug flow jets

G. Gabard[a)]
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 27 March 2008; revised 22 July 2008; accepted 26 July 2008)

This paper reports a theoretical study of the radiation of sound through jet exhausts. It focuses on the transition from near field to far field by considering the features of the near-field solution and how these features translate to the far field. The main focus of this work is the importance in some cases of lateral waves radiating from the jet. While the presence of lateral waves has long been recognized, there has been no systematic investigation of the practical consequences of these waves in the prediction of sound propagation through round jets. The physical mechanisms involved in the generation of these waves are presented as well as the conditions under which they become significant. Another issue is the possibility of "channeled waves" inside the jet associated with strong sound radiation in the forward arc. This paper also discusses the validity of the far-field approximation when lateral waves are present. It is shown that the standard far-field approximation can be improved by adding correction terms that account for the presence of the lateral waves and channeled waves. The challenge posed to computational aeroacoustics by these near-field effects is also discussed. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977609]

PACS number(s): 43.28.Py, 43.20.El [JWP] Pages: 2755–2766

I. INTRODUCTION

Predicting the propagation of sound through jets is of particular importance for jet noise modeling and for fan exhaust noise simulations. For both applications it is crucial to account for the refraction of sound through the mixing layer of the jet. For jet noise modeling, a simplified description corresponding to a low frequency approximation of the real problem is that of a point source embedded in a plug flow jet where the mixing layer is described as a cylindrical discontinuity in the mean flow.[1,2] For fan exhaust noise, a classical idealized model is the Munt solution, which involves the radiation of sound from a semi-infinite jet pipe carrying a plug flow jet.[3]

These two model problems will be used in the present paper to investigate the transition of the radiated sound from the near field to the far field. In the process the validity of the far-field approximation will be discussed and defined more precisely. While noise radiation in classical acoustics with no flow can be understood relatively easily, for instance, by means of ray acoustics, the jet introduces additional features in the propagation of sound to the far field. The two features to be discussed in detail are the lateral waves radiated downstream of the source and the channeled waves. Although these features are essentially associated with near-field effects due to the vortex sheet, they can influence the radiated sound at large distances from the source, that is, at distances where the far-field approximation would usually apply. The present paper also aims at providing a description of the conditions required for these effects to be important in the context of the plug flow jet so that analytical models such as the Munt solution can be used with more confidence. In fact the present work is a continuation of that reported in Ref. 4 where numerical simulations of the Munt problem were compared to the far-field analytic solution, and unexpected discrepancies were observed.

The validity of the far-field approximation is not only a theoretical issue but is also of importance in jet noise or turbomachinery noise measurements where it is important to have microphones located in the geometric far field in order to use the spherical spreading to extrapolate the measurements for larger distances. Measurements can be influenced by near-field effects and nonlinearities if the microphones are too close to the jet.[5,6]

The key of the problem is the transmission and reflection of sound at the vortex sheet separating the jet from the ambient flow. Several theoretical analyses were devoted to the interaction of linear acoustic disturbances with a vortex sheet. Earlier work focused on the continuity conditions to be imposed at the vortex sheet, that is, the continuity of pressure and normal displacement. A summary of this work is given by Howe.[7]

A central issue is the stability of the vortex sheet, which exhibits a linear Kelvin–Helmholtz instability. It induces an acoustic field growing exponentially in the streamwise direction and thus dominates the sound radiated from the acoustic source. The unlimited exponential growth of this linear instability is nonphysical, and its relevance for the sound radiated to the far field is therefore questionable. However, as first noted by Miles,[8] the instability wave should be included in order to obtain a causal solution. The formulation of causal solutions has since been further investigated by Jones,[9,10] Jones and Morgan,[11] and by Crighton and Leppington.[12]

Another important aspect, and the topic of this paper, is the presence of lateral waves that can be particularly signifi-

[a)]Electronic mail: gabard@soton.ac.uk

cant inside the zone of silence. Lateral waves represent a distinct part of the scattered acoustic field in addition to the direct reflected and transmitted waves. These were first included in the analysis by Gottlieb who derived correction terms for the far-field approximation of the solution.[13] For a planar vortex sheet, the derivation of a complete solution taking both the instability and the lateral waves into account was initiated by Friedland and Pierce[14] and completed by Jones and Morgan.[11]

For axisymmetric jets, Morgan derived the exact solution for a fixed monopole embedded in a subsonic plug flow jet.[1] However, lateral waves were not included in the analysis as it was incorrectly assumed that their behavior was similar to that of the two-dimensional vortex sheet. Goldstein also derived solutions in the low frequency limit for convected point sources in an axisymmetric jet,[15,16] but lateral waves were not considered. The present paper aims at providing a detailed account of near-field effects, including lateral waves, in the context of round jets using the plug flow as a simplified model for the shear layer.

It should be noted that noise radiation from round jets has been extensively studied in the high frequency limit (see, for instance, Ref. 17). However, again lateral waves have seldom been considered. For planar vortex sheets Suzuki and Lele studied both the low and high frequency limits for the lateral waves and compared these results with direct numerical simulations.[18,19]

The remainder of this paper is as follows. The plug flow jet and the Munt problem are introduced in the next section together with their analytical solutions. The standard far-field approximation and a uniformly valid approximation are discussed in Sec. III. In Sec. IV, the presence of lateral waves is illustrated and analyzed. The properties of channeled waves are described in Sec. V.

II. DESCRIPTION OF THE PROBLEMS

The present analysis is concerned with two model problems in aeroacoustics shown in Fig. 1. The first one is the radiation of sound by a fixed point source in a plug flow jet. The second one is the Munt problem, which involves the radiation of sound from a semi-infinite jet pipe carrying a plug flow jet.

A. The plug flow jet

This is an idealized model for the propagation of sound through a jet mixing layer, and it forms the basis for several jet noise models.[2] The jet is a cylinder with radius R_0 where the flow is purely axial and uniform with velocity v_j, density ρ_j, and sound speed c_j. The jet is embedded in a uniform axial ambient flow with velocity v_0, density ρ_0, and sound speed c_0. The continuity of pressure and normal displacement apply across the vortex sheet separating the jet and the ambient flow. An important aspect of the problem is the refraction of sound by the flow discontinuity across the vortex sheet, which results in the so-called cone of silence. This model represents a low frequency approximation of sound propagation through jets in the sense that the mixing layer is small compared to the acoustic wavelength. The model is presented in nondimensional form with R_0, c_0, and ρ_0 as reference values.

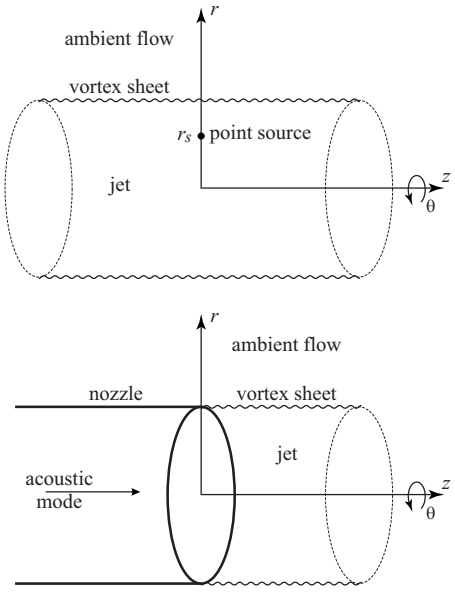


FIG. 1. Top: plug flow jet with a point source. Bottom: the Munt problem.

For a source $Q=\delta(r-r_s)\delta(z)e^{i\omega t-im\theta}/r_s$ representing a circumferential Fourier component of a time-harmonic mass point source located inside the jet ($r_s<1$), the complex amplitude of the pressure field outside the jet can be written as an inverse Fourier transform,

$$p(r,z)=\frac{\omega}{2\pi}\int_\Gamma P(u)H_m^{(2)}(\lambda_0\omega r)e^{-i\omega uz}du, \tag{1}$$

where $H_m^{(2)}$ is the Hankel function of the second type and u denotes the reduced axial wavenumber (ωu is the actual axial wavenumber). The superscript (2) for the Hankel function will be omitted in what follows. The solution is given by

$$P(u)=i(1-uM_0)^2(1-uM_1)\frac{J_m(\lambda_1\omega r_s)}{L(u)}, \tag{2}$$

where $M_0=v_0/c_0$, $M_1=v_j/c_0$, and J_m is the Bessel function. The radial wavenumbers in the jet and the ambient flow are $\lambda_1(u)=[C_1^2(1-uM_1)^2-u^2]^{1/2}$ and $\lambda_0(u)=[(1-uM_0)^2-u^2]^{1/2}$ with $C_1=c_0/c_j$. The branch cuts of λ_0 and λ_1 are defined from $u_{0,1}^{\pm}$ to $u_{0,1}^{\pm}\mp i\infty$, with $u_0^{\pm}=1/(M_0\pm1)$ and $u_1^{\pm}=C_1/(C_1M_1\pm1)$.

The function $L(u)$ is of particular importance as $L(u)=0$ represents the characteristic equation of the jet, which specifies the modes of oscillation of the jet column. It is defined as

$$L(u)=D_1(1-uM_1)^2J_m(\lambda_1\omega)\lambda_0H_m'(\lambda_0\omega)-(1-uM_0)^2H_m(\lambda_0\omega)\lambda_1J_m'(\lambda_1\omega),$$

where $D_1=\rho_j/\rho_0$. One mode of the jet corresponds to the Kelvin–Helmholtz instability of the vortex sheet. It is located at $u=u_0$ in the first quadrant of the complex plane and represents a disturbance growing exponentially in the axial di-

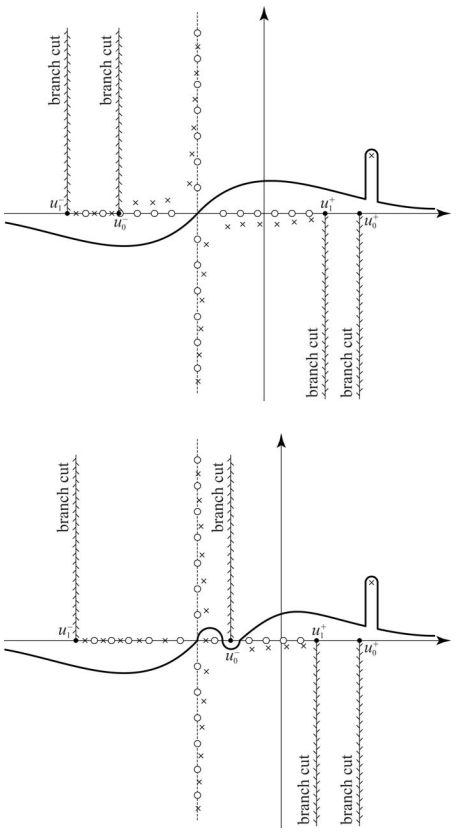


FIG. 2. Locations of the zeros of $L(u)$ (×) and the zeros of $J'_m(\lambda_1\omega)$ (○) in the complex u plane. Top: with $M_0 > 1-M_1C_1^2/(1-M_1^2C_1^2)$; bottom: with $M_0 < 1-M_1C_1^2/(1-M_1^2C_1^2)$. The dashed line is $\mathrm{Re}(u)=u_c$. The thick line is the integration contour Γ. In the bottom figure note the deformation of the integration contour around the branch point u_0^-.

rection. A high frequency approximation yields the following estimate:[20]

$$u_0^{-1} \simeq \frac{M_0+M_1}{2} + \frac{i}{2}[1-(\sqrt{1+(M_1-M_0)^2}-2)^2]^{1/2}.$$

In addition to the instability mode, there is a series of acoustic modes corresponding to exponentially decaying oscillations of the jet. These modes are located either close to the real axis between u_1^- and u_1^+ or close to the line $\mathrm{Re}(u)=u_c=(u_1^-+u_1^+)/2$ (see Fig. 2). The zeros located in the region $-\pi/2 \leqslant \arg(u-u_c) \leqslant 0$ represent modes propagating in the positive axial direction, while the zeros in the region $\pi/2 \leqslant \arg(u-u_c) \leqslant \pi$ represent modes propagating in the negative direction. The physical relevance of the location of these zeros will be discussed in more detail in Secs. IV and V. Finally, there are two series of zeros clustered around u_0^+ and u_0^-, as shown in Fig. 3 (in the lower and upper parts of the complex plane, respectively).

The contour of integration Γ for the inverse Fourier transform is also shown in Fig. 2. This contour is defined from $-\infty-i0$ to $+\infty+i0$, and it intersects the real axis at u_c. In order to obtain a causal solution, it is necessary to deform the contour around the point u_0 corresponding to the instability wave of the vortex sheet.[12]

For sufficiently high jet Mach numbers, the branch point u_0^- can move to the right of u_c [this is the case when $M_0 < 1-M_1C_1^2/(1-M_1^2C_1^2)$]. In that situation the physics of the problem is slightly modified since the axial phase speed of a downstream mode of the jet can become subsonic with respect to the ambient medium. The consequence of this will be discussed in detail in Sec. V. As noted by Morgan,[1] another consequence is that one has to deform the contour of integration so as to cross the real axis at u_c and then to pass below the branch point u_0^-, as illustrated in Fig. 2.

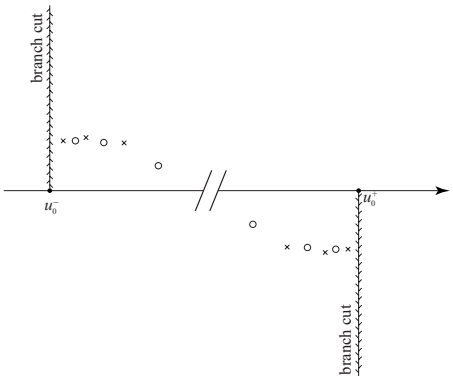


FIG. 3. Close-up view of the branch points u_0^- and u_0^+ with the zeros of $L(u)$ (×) and the zeros of $H'_m(\lambda_0\omega)$ (○). The parameters are $\omega=20$, $m=18$, $M_1=0.5$, $M_0=0$, and $C_1=D_1=1$.

B. The Munt problem

The Munt problem is an extension of the problem presented above where the jet exhaust is represented by a semi-infinite duct. The source of sound is an incident mode propagating from inside the duct toward the duct exhaust. This acoustic mode radiates out of the duct through the mixing layer and to the far field. This is a simplified description of the sound radiation from turbofan exhausts. An analytical solution for this problem was first derived by Munt using the Wiener–Hopf technique.[3] Several extensions of the original Munt problem have been recently developed.[20–22] In Ref. 20, it was shown that it is possible to evaluate exactly the analytical solution in the near field (that is, without the far-field approximation). The pressure field of the incident acoustic mode is $p(r,z,\theta)=\Psi_{mn}(r)e^{-i\omega\mu_{mn}^+z-im\theta}$, where $\Psi_{mn}(r)$ is the mode shape, μ_{mn}^+ is the axial wavenumber, and m and n are the azimuthal and radial mode orders. It is possible to obtain either the noncausal solution, which ignores the instability wave of the vortex sheet, or the causal solution, which includes the instability.[12,23] In the latter case it is also possible to separate explicitly the instability from the rest of the solution, and thus the causal acoustic field alone can be obtained. When the full Kutta condition is applied at the duct lip, the solution for the Munt problem is given by[20]

$$P(u) = i\Psi_{mn}(1)\lambda_0^-(\mu_{mn}^+)\lambda_1^-(\mu_{mn}^+)K_-(\mu_{mn}^+) \times \frac{(1-uM_0)^2\lambda_1^+(u)}{\omega(u-\mu_{mn}^+)\lambda_0^-(u)K_+(u)H'_m(\lambda_0\omega)}, \quad (3)$$

where $\lambda_0^\pm=[1-(M_0\pm1)u]^{1/2}$ and $\lambda_0^\pm=[C_1-(M_1C_1\pm1)u]^{1/2}$ (the branch cuts are the same as for λ_0 and λ_1). The addition of the semi-infinite duct introduces a major difficulty since one has to find two split functions, $K_+(u)$ and $K_-(u)$, regular and nonzero in the upper and lower halves of the complex u

plane, respectively, such that $K(u)=K_+(u)/K_-(u)$. This factorization of the kernel is not trivial but will not be described here (see Ref. 24). The Wiener–Hopf kernel is defined by $K(u)=L(u)/J'_m(\lambda_1\omega)/H'_m(\lambda_0\omega)$, and it has the same zeros as the function $L(u)$ for the jet column presented above. In addition, the poles of $K(u)$ corresponds either to zeros of $J'_m(\lambda_1\omega)$ or to zeros of $H'_m(\lambda_0\omega)$. The first case represents acoustic modes in an infinite straight circular duct. As illustrated in Fig. 2, these poles are located either on the real axis between u_1^- and u_1^+ (for cut-on modes) or on the line $\mathrm{Re}(u)=u_c$ (for cut-off modes). The zeros of $H'_m(\lambda_0\omega)$ represent solutions to the exterior problem, that is, the diffraction of sound by an infinite rigid cylinder embedded in the ambient flow. These poles are clustered near u_0^- and u_0^+, as shown in Fig. 3 (see p. 373 of Ref. 25 for a description of the zeros of H_m).

III. FAR-FIELD ASYMPTOTIC SOLUTION

Although it is possible to evaluate the exact solutions of the plug flow jet and of the Munt problem in the near field using Eq. (1), it is common practice to use the far-field approximation of these solutions. It is justified by the fact that the observer is generally many wavelengths away from the source embedded in the jet or from the duct exhaust. The advantage of the far-field approximation is that one does not need to evaluate the complex integral for the inverse Fourier transform. Also the far-field approximation provides a clear distinction between the geometrical spreading and the directivity of the acoustic field. In preparation for the analysis of the transition from near to far field, the derivation of the standard far-field approximation is reviewed in this section, and a uniformly valid approximation is also derived.

A. The standard far-field approximation

For the present analysis it is necessary to make explicit the behavior at infinity of the Hankel function in the integrand of the inverse Fourier transform [Eq. (1)]. To that end we consider

$$p(r,z)=\frac{\omega i^{m+1/2}}{\sqrt{2\pi^3}}\int_\Gamma \frac{P(u)Q(\lambda_0\omega r)}{\sqrt{\lambda_0\omega r}}e^{-i\omega uz-i\lambda_0\omega r}du, \tag{4}$$

which is an exact reformulation of Eq. (1). The additional function Q in the integrand is the ratio between the Hankel function and its asymptotic expression for a large argument (see item 9.2.3 in Ref. 25),

$$Q(z)=\left(\frac{\pi z}{2}\right)^{1/2}H_m(z)e^{iz-im\pi/2-i\pi/4}.$$

When $z\gg|m^2-1/4|$, we have $Q(z)\sim 1$. This approximation is almost systematically used in previous work but is not mandatory.

We now introduce the following change of variables corresponding to a Lorentz transform that converts the ambient flow into a steady medium:

$$r=\hat{R}\sin\hat{\varphi},\quad z=\sqrt{1-M_0^2}\hat{R}\cos\hat{\varphi},$$

$$\omega=\sqrt{1-M_0^2}\hat{\omega},\quad u=\frac{\tau-M_0}{1-M_0^2}.$$

The solution written with this Lorentz transform is the acoustic field perceived by an observer moving with the ambient flow. Equation (4) now reads

$$p(\hat{R},\hat{\varphi})=\frac{\hat{\omega}i^{m+1/2}e^{i\hat{\omega}\hat{R}M_0\cos\hat{\varphi}}}{[2\pi^3(1-M_0^2)\hat{\omega}\hat{R}\sin\hat{\varphi}]^{1/2}}\times\int_\Gamma\frac{P(u(\tau))Q(\hat{\omega}\hat{R}\sqrt{1-\tau^2}\sin\hat{\varphi})}{(1-\tau^2)^{1/4}}e^{\hat{\omega}\hat{R}q(\tau)}d\tau, \tag{5}$$

with $q(\tau)=-i\sqrt{1-r^2}\sin\hat{\varphi}-i\tau\cos\hat{\varphi}$.

When $\hat{\omega}\hat{R}\to\infty$, the complex integral in Eq. (5) can be approximated using the method of steepest descent (see Chap. 4 of Ref. 26). The function q is analytic except for the branch cuts of the square root, which correspond to the branch cuts of λ_0 and go from ±1 to $\pm1\mp i\infty$ in the τ complex plane. It has a single first-order saddle point at $\tau_s=\cos\hat{\varphi}$, where we have $q(\tau_s)=-i$ and $q''(\tau_s)=i/\sin(\hat{\varphi})^2$. From that saddle point, a path of steepest descent can be defined by $q(\tau)=q(\tau_s)-s^2$, where s is a real parameter. This yields the following parametrization of the path of steepest descent:

$$\tau=\cos(\hat{\varphi}\mp\cos^{-1}(1-is^2))\quad\text{with } s\gtrless 0.$$

The contour Γ can be deformed onto the path of steepest descent provided that any poles or branch cuts crossed in the process are properly accounted for. We will deal in detail with this aspect in the next section. On the path of steepest descent, the integrand decays exponentially away from the saddle point, and only the contribution from that point is significant. The integral in Eq. (5) can then be approximated by

$$\sqrt{\frac{-2\pi}{\hat{\omega}\hat{R}q''(\tau_s)}}\frac{P(u_s)Q(\hat{\omega}\hat{R}\sin^2\hat{\varphi})}{(1-\tau_s^2)^{1/4}}e^{\hat{\omega}\hat{R}q(\tau_s)}, \tag{6}$$

where $u_s=u(\tau_s)$ and the argument of the square root is taken to be $\arg(d\tau)_{\tau_s}=\pi/4$, with $d\tau$ denoting an element of the steepest descent path at τ_s (see p. 382 of Ref. 26). The asymptotic far-field solution is given by

$$p(\hat{R},\hat{\varphi})\sim\frac{i^{m+1}P(u_s)Q(\hat{\omega}\hat{R}\sin^2\hat{\varphi})}{\pi\hat{R}\sqrt{1-M_0^2}}e^{-i\hat{\omega}\hat{R}(1-M_0\cos\hat{\varphi})}. \tag{7}$$

We now apply the inverse Lorentz transform by introducing the spherical coordinates R-φ in the physical frame of reference defined by $z=R\cos\varphi$ and $r=R\sin\varphi$. This yields

$$p(R,\varphi)\sim\frac{i^{m+1}P(u_s)e^{-i\omega RS(\varphi)}}{\pi R\sqrt{1-M_0^2\sin^2\varphi}}Q\left[\frac{\omega R\sin^2\varphi}{(1-M_0^2\sin^2\varphi)^{1/2}}\right], \tag{8}$$

where $S(\varphi)=[(1-M_0^2\sin^2\varphi)^{1/2}-M_0\cos\varphi]/(1-M_0^2)$ accounts for the Doppler effect of the ambient flow. The standard spherical spreading of the acoustic waves in the far field is not yet apparent in Eq. (8) due to the dependence of the

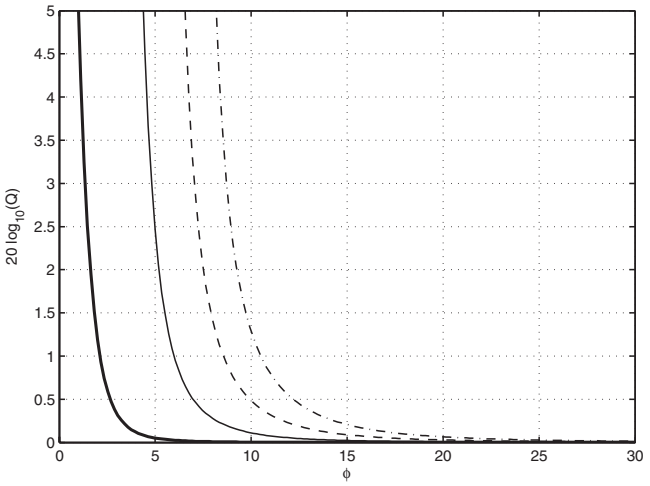


FIG. 4. Amplitude of the factor Q in dB with $\omega R=80$ and $M_0=0$. Thick solid line: $m=1$; thin solid line: $m=5$; dashed line: $m=10$; dot-dashed line: $m=15$.

function Q on R. To recover the standard $1/R$ spherical spreading, one must assume that $Q\approx 1$, which is valid when $R \gg |m^2-1/4|(1-M_0^2 \sin^2\varphi)^{1/2}/(\omega \sin^2\varphi)$. Note that the factor Q does not depend on the details of the jet and the source. Figure 4 shows the factor Q as a function of φ for various azimuthal orders m. It is clear that the approximation $Q\approx 1$ is not valid close to the jet axis, especially for high azimuthal orders, since Q increases very rapidly as φ decreases. However, it should be mentioned that for large azimuthal orders, sound is primarily radiated at large angles so the approximation $Q\approx 1$ should not impair the prediction of the main peak of directivity. Although not shown in Fig. 4, it was found that the influence of M_0 on Q is negligible for small angle φ.

B. A uniformly valid approximation

The standard asymptotic far field [Eq. (8)] is only valid if there is no pole in the integrand close to the path of steepest descent and if no pole has been crossed when the contour Γ is deformed onto the path of steepest descent. This requirement is generally not satisfied, and the standard approximation [Eq. (8)] is not uniformly valid over the range of direction φ.

Consider that $P(u)$ has a pole close to the path of steepest descent, say, at $\tau=\tau_p$. The contribution of the integrand in the vicinity of the pole is not negligible compared to the contribution of the saddle point. Therefore, the following term should be added to approximation (6) of the integral (see Sec. 4.4a of Ref. 26):

$$\frac{\sqrt{\pi}\hat{A}_p Q(\hat{\omega}\hat{R}\sqrt{1-\tau_p^2}\sin\hat{\varphi})}{s_p\sqrt{\hat{\omega}\hat{R}}(1-\tau_p^2)^{1/4}} e^{\hat{\omega}\hat{R}_{q(\tau_s)}} C(s_p\sqrt{\hat{\omega}\hat{R}}),$$

with

$$C(z)=1\pm i\sqrt{\pi}w(\pm z)z \quad \text{for } \mathrm{Im}(z)\gtrless 0,$$

where w is the error function defined by item 7.1.3 of Ref. 25. $\hat{A}_p$ is the residue of $P(u(\tau))$ at $\tau=\tau_p$. The position of the pole in the complex s plane is $s_p=\sqrt{q(\tau_s)-q(\tau_p)}$, where the branch of the square root is chosen so that $s_p\to(\tau_p-\tau_s)/\sqrt{-2/q''(\tau_s)}$ as $\tau_p\to\tau_s$ [the argument of the square root in this last expression is $\arg(d\tau)_{\tau_s}=\pi/4$]. This yields a first correction term f_p that needs to be added to the standard approximation [Eq. (7)],

$$f_p(\hat{R},\hat{\varphi})=\frac{i^{m+1/2}\hat{A}_p e^{-i\hat{\omega}\hat{R}(1-M_0\cos\hat{\varphi})}}{\pi s_p\hat{R}(1-\tau_p^2)^{1/4}[2(1-M_0^2)\sin\hat{\varphi}]^{1/2}} \times C(s_p\sqrt{\hat{\omega}\hat{R}})Q(\hat{\omega}\hat{R}\sqrt{1-\tau_p^2}\sin\hat{\varphi}).$$

This expression can also be written with the physical spherical coordinates R and φ,

$$f_p(R,\varphi)=\frac{i^{m+1/2}\sqrt{1-M_0^2}A_p e^{-i\omega RS(\varphi)}}{\pi s_p R\sqrt{2\lambda_0(u_0)\sin\varphi}(1-M_0^2\sin^2\varphi)^{1/4}} \times C(s_p\sqrt{\hat{\omega}\hat{R}})Q(\lambda_0(u_p)\omega r), \quad (9)$$

where A_p is the residue of $P(u)$ at $u=u_p$. This term becomes significant when $|q(\tau_s)-q(\tau_p)|^{-1}\sim O(\hat{\omega}\hat{R})$, and it plays a similar role to the transition functions near shadow region boundaries in ray acoustics (see p. 405 of Ref. 26).

The correction term described above accounts for the presence of a pole in the vicinity of the path of steepest descent. Another term should be added in the asymptotic expression if a pole is crossed when the contour of integration is deformed onto the path of steepest descent. This term corresponds to the residue of the integrand in Eq. (5),

$$g_p(\hat{R},\hat{\varphi})=\pm\frac{i^{m+3/2}\sqrt{2\hat{\omega}}\hat{A}_p e^{\hat{\omega}\hat{R}(iM_0\cos\hat{\varphi}+q(\tau_p))}}{[\pi(1-M_0^2)\hat{R}\sin\hat{\varphi}]^{1/2}(1-\tau_p^2)^{1/4}} \times Q(\hat{\omega}\hat{R}\sqrt{1-\tau_p^2}\sin\hat{\varphi})H(\mathrm{Im}(s_p)),$$

where H is the Heaviside function. The sign in this expression depends on the position of the pole relative to the path of steepest descent. With the physical spherical coordinates this expression reads

$$g_p(R,\varphi)=\pm i^{m+3/2}A_p\frac{\sqrt{2\omega}e^{-i\omega RS(\varphi)-\hat{\omega}\hat{R}s_p^2}}{\sqrt{\pi r\lambda_0(u_p)}} \times Q(\lambda_0(u_p)\omega r)H(\mathrm{Im}(s_p)). \quad (10)$$

For a series of poles $\{u_p\}_{p=1,\ldots,N}$, the uniformly valid asymptotic solution in the far field is given by

$$p(R,\varphi)\sim\frac{i^{m+1}P(u_s)e^{-i\omega RS(\varphi)}}{\pi R\sqrt{1-M_0^2\sin^2\varphi}}Q\left[\frac{\omega R\sin^2\varphi}{(1-M_0^2\sin^2\varphi)^{1/2}}\right] + \sum_{p=1}^{N}[f_p(R,\varphi)+g_p(R,\varphi)]. \quad (11)$$

Although the two terms f_p and g_p are discontinuous when the sign of $\mathrm{Im}(s_p)$ changes, the sum f_p+g_p is continuous. Jones and Morgan derived similar correction terms for a two-dimensional problem with a planar vortex sheet.[11]

C. Instability

Note that in Eq. (8) the Kelvin–Helmholtz instability of the vortex sheet is not accounted for, so Eq. (8) represents

the far-field approximation of the acoustic solution alone. The instability wave is associated with the pole of the integrand located in the first quadrant of the complex u plane, as shown in Fig. 2. The corresponding pressure field is readily obtained by taking its residue in Eq. (1),

$$p(r,z) = i\omega A_0 H_m(\lambda_0(u_0)\omega r) e^{-i\omega u_0 z} H(z),$$

where A_0 is the residue of $P(u)$ at u_0. This pressure field is exponentially growing in a conical region defined by $\varphi < \varphi^\star$ with $\tan \varphi^* = -\mathrm{Im}(u_0)/\mathrm{Im}[\lambda_0(u_0)]$. Outside this region it is exponentially decaying and negligible.[20]

IV. LATERAL WAVES

The transition of the acoustic solutions from the near field to the far field is now discussed. The first feature is the presence of lateral waves.

A. Observations and physical principles

Results are presented here for the Munt problem with a circular duct carrying a jet with Mach number $v_j/c_j = 0.447$, sound speed $c_j/c_0 = 1.02$, and density $\rho_j/\rho_0 = 0.96$. The ambient medium is steady. The Helmholtz number based on the duct radius is 19.4. This configuration is representative of an aeroengine bypass duct at approach conditions. To study the transition from the near-field solution to the standard far-field approximation, the pressure is plotted along a circle centered on the exhaust plane. The radius of the circle is then varied from $R = 3.3$ to 38 (the latter corresponding approximately to the distance used for engine certification measurements). Figure 5 shows the results obtained with the analytical model for an incident plane wave. With the far-field solution, the cone of silence is approximately 47° wide. The directivity then reaches a sharp peak followed by a series of lobes of directivity. For the near-field solutions outside the cone of silence, the pattern of the far-field approximation quickly emerges as the distance of the observer is increased, with the lobes and zeros of the solutions rapidly approaching that of the far-field solution. It can therefore be concluded that the far-field asymptotic solution is a good approximation of the exact solution in this range of directions. However, two important differences can be observed between the far-field solution and the series of near-field solutions.

Firstly inside the cone of silence, as the distance is increased, the pressure amplitude does not converge to the far-field asymptotic solution (at least for distances up to $R = 38$, which corresponds in this case to more than 117 wavelengths) and remains comparable in amplitude to the main lobe of directivity at around 90 dB. This observation suggests that there is a feature of the sound radiated in the cone of silence in the near field that is not captured by the far-field approximation. Also the fact that the far-field asymptotic solution is not recovered even at 117 wavelengths from the duct exhaust indicates that the common definition of the far field does not apply in that case.

Secondly, there are significant differences for the main lobe of directivity both in terms of direction and amplitude. While a sharp peak at 48° with an amplitude of 105 dB is observed in the far-field solution, the main lobe in the near field is smooth, located at approximately 46° and with an amplitude of 93 dB. In fact the sound amplitude observed in the near field of the cone of silence is almost as high as the main lobe of directivity. These differences are particularly significant since the direction and amplitude of the main lobe are often the main information to be obtained from fan noise simulations, and they also have a significant impact on the predictions of perceived noise levels.

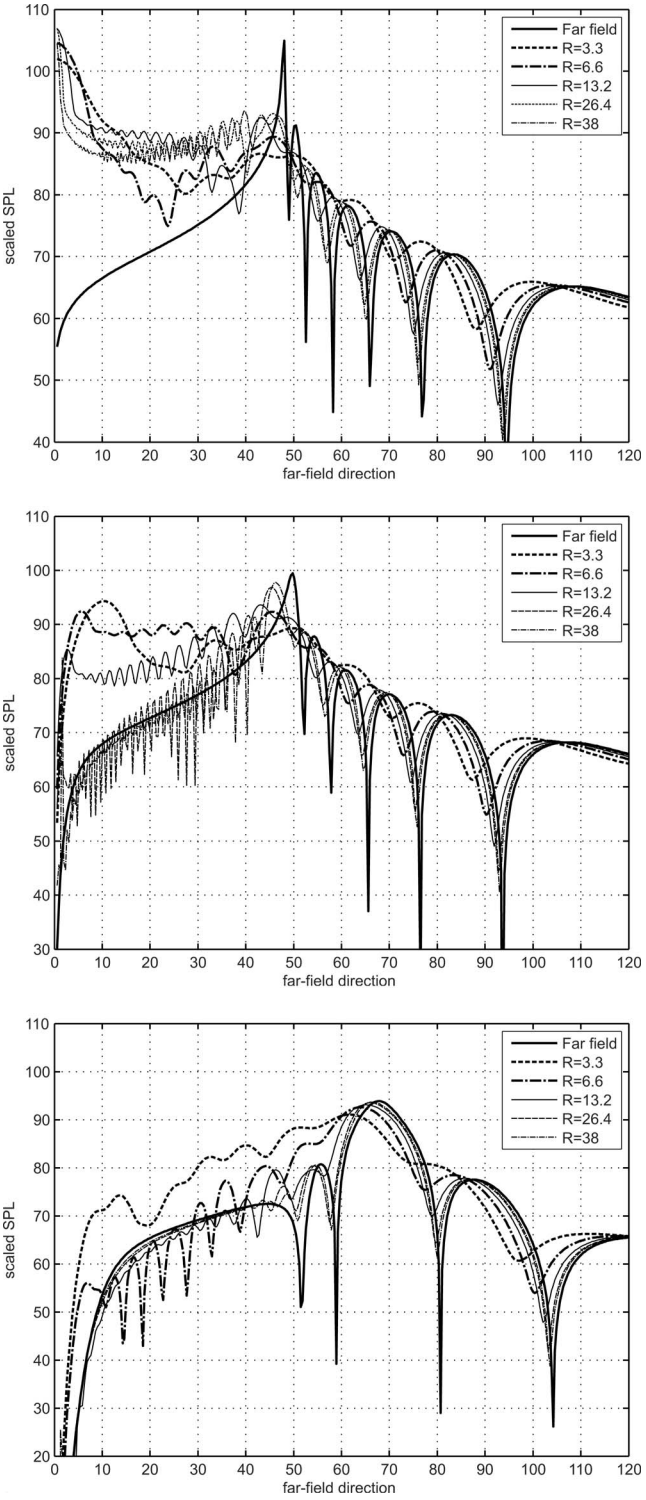


FIG. 5. Transition from near to far field for the Munt problem with an incident plane wave (top) with mode (2,0) (center) and mode (5,2) (bottom). All the results have been scaled for $R = 38$.

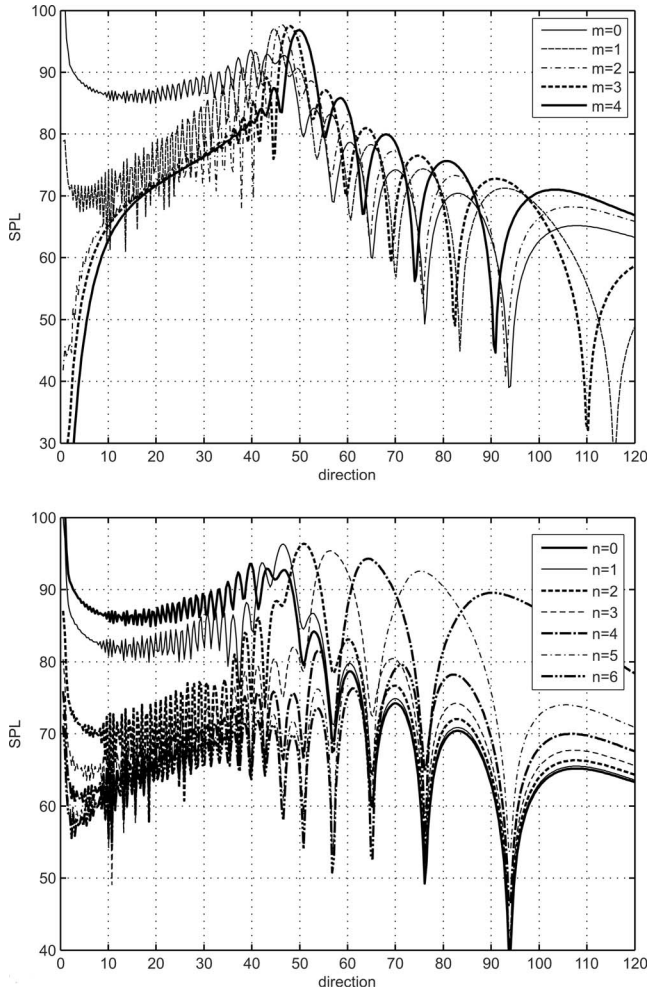


FIG. 6. Exact near-field solutions at $R=38$. Left: first radial mode with varying azimuthal order. Right: axisymmetric modes ($m=0$) with varying radial order.

The agreement between the near-field and far-field solutions tends, however, to improve for higher-order modes. Two examples are given in Fig. 5 with the modes (2,0) and (5,2). While there are still significant differences at short distances from the exhaust, the near-field solution eventually collapses onto the far-field approximation. This is also illustrated in a different way in Fig. 6 where the directivity at a fixed distance $R=38$ is plotted for various azimuthal and radial orders. While the lateral waves are more important for the plane wave, their amplitudes remain significant for higher azimuthal and radial orders.

The sound field observed in the near field inside the cone of silence is due to the presence of lateral waves. They are found in problems of wave propagation across interfaces between media with different propagation speeds, such as underwater acoustics, linear elasticity, and electromagnetism (see Sec. 30 of Ref. 27 and see also Ref. 28). In the context of aeroacoustics, the presence of lateral waves has long been recognized in the theoretical analysis of an acoustic wave interacting with a planar vortex sheet.[11–14] More recently Suzuki and Lele investigated numerically and analytically the properties of refracted arrival waves in planar mixing layers.[18,19] However, the practical consequences of these waves for the prediction of noise radiation from round jets have not yet been studied.

In the context of the Munt problem, the physical mechanism involved in the presence of lateral waves is the following. When well cut-on acoustic modes radiate from the duct exhaust (that is, low-order modes at relatively high Helmholtz numbers), the angle of incidence of the waves on the vortex sheet is small. This results in a weak transmission of acoustic energy through the vortex sheet into the ambient flow (diffraction theory shows that transmission decreases and then vanishes for shallow angles of incidence). Hence, the sound waves are partly confined inside the jet due to significant reflections at the vortex sheet. In that case, acoustic waves are allowed to propagate far away from the duct exhaust within the jet and do not decay as fast as the standard decay rate. This is clearly visible in Fig. 7 where the pressure amplitude in the jet is still significant 16 wavelengths away from the exhaust. This is also shown in Fig. 5 where the sound pressure level inside the jet is visible close to the axis and is actually higher than that in the ambient flow by more than 10 dB. A consequence is that the lateral waves in the ambient flow are generated continuously along a large portion of the vortex sheet, which can extend far away from the duct exhaust. In contrast, the sound corresponding to the typical lobes of directivity observed in the far field radiates from a small region near the exhaust. The generation of lateral waves for the point source in a plug flow jet can be understood in a similar way.

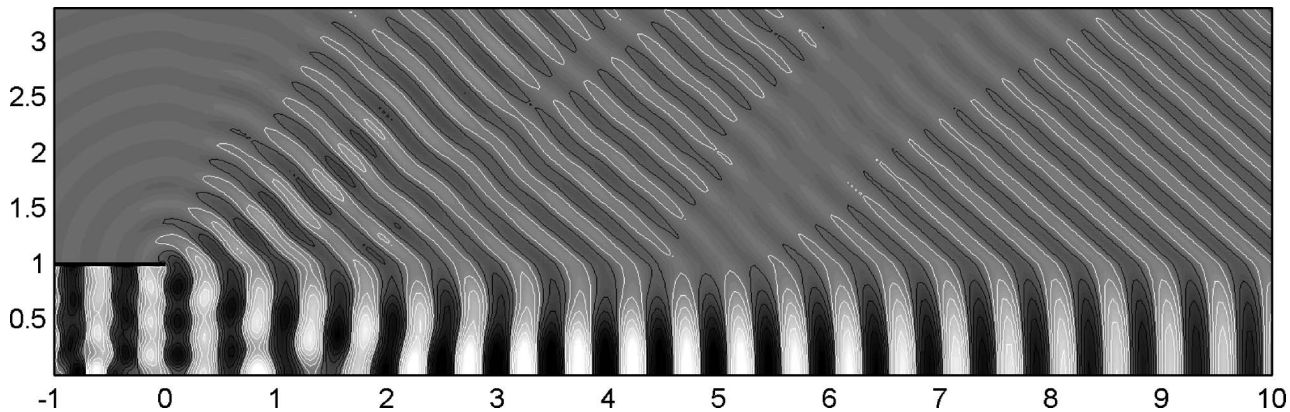


FIG. 7. Real part of pressure in the near field for the Munt problem with the plane wave.

B. Analysis

In theoretical studies of sound interacting with planar vortex sheets, the lateral waves are associated with a branch cut in the integrand of the inverse Fourier transform [Eq. (1)]. When the contour is deformed onto the steepest descent path, it is indented around the branch cut, and it is this contribution of the contour that accounts for the lateral waves. This branch cut is associated with the transverse wavenumber λ_1 in the region with a faster propagation speed. As indicated in Sec. I, previous work on lateral waves focused on planar vortex sheets, and it turns out that the situation is different for cylindrical vortex sheets. As shown in Appendix A, the branch cuts of λ_1 are not present in the integrands. So apart from the instability wave, the path of steepest descent does not need to be indented. The lateral waves are described by the analytical solutions in a different way. In fact their presence can be traced back to the contribution of the fastest downstream modes of the jet. These modes correspond to the zeros of $L(u)$ located close to u_1^+. As can be seen in Fig. 2, these zeros can be close to the real axis. When this is the case, these modes only slowly decay in the axial direction and, if excited, will generate significant oscillations of the vortex sheet far away from the duct exhaust.

It is possible to obtain an approximation for the location of these zeros in the complex u plane,

$$\mu_{mn}^{+} - \frac{D_1(1-\mu_{mn}^{+}M_1)^2}{\omega(1-\mu_{mn}^{+}M_0)^2}\frac{\lambda_0(\mu_{mn}^{+})H_m'(\lambda_0(\mu_{mn}^{+})\omega)}{H_m(\lambda_0(\mu_{mn}^{+})\omega)} \times \frac{J_m(j_{mn}')}{J_m''(j_{mn}')}[M_1C_1^2+(1-M_1^2C_1^2)u]^{-1}, \quad (12)$$

where j_{mn}' denotes the zeros of $J_m'(z)$, so we have $\lambda_1(\mu_{mn}^{+})\omega = j_{mn}'$. Details are given in Appendix B. Although this expression provides only a qualitative estimate of the location of the zeros, it is interesting to note that the zeros move closer to the real axis as the frequency increases or as the relative density of the jet decreases.

The presence of these zeros close to the real axis renders the standard far-field approximation (8) inaccurate. As explained in Sec. III B, Eq. (8) is valid as long as no pole is located in the vicinity of the steepest descent path. As the far-field angle φ is reduced from $\pi/2$ to 0, the saddle point moves closer to the pole, and then for some critical angle φ_c the pole is located right on the steepest descent path. For φ close to or smaller than φ_c, it is necessary to use the uniformly valid approximation (11), which accounts for the presence of this pole. An example of results obtained with the modified asymptotic solution is given in Fig. 8, where the contributions of the first and second fastest modes of the jet are taken into account. The correction terms have two effects. First, the sharp peak observed in the standard far-field solution is significantly reduced. Second, the lateral waves present in the cone of silence are now included in the far-field approximation. This demonstrates that their presence is controlled by the amplitude of these modes of the jet. The time required to compute the correction terms for the far-field solution is negligible, but the accuracy of the asymptotic solution is significantly improved.

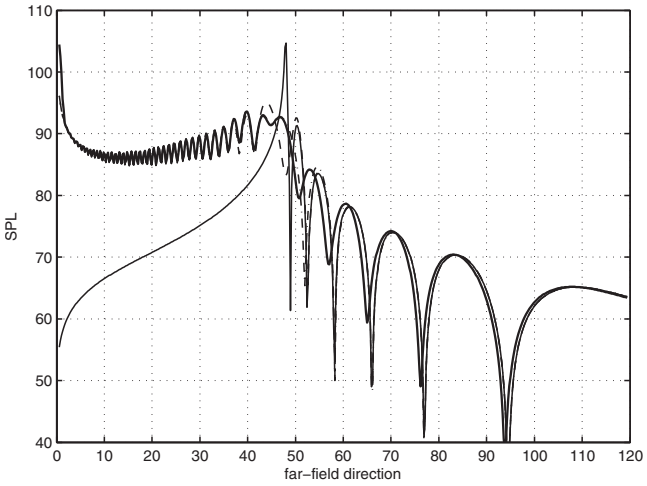


FIG. 8. Comparison of the standard asymptotic far field (thin solid line), with the modified asymptotic solutions including the contribution from the fastest jet mode (dashed line) and the contributions from the fastest two jet modes (dot-dashed line). The near-field solution is also plotted (thick solid line). All solutions are for $R=38$.

In order to provide a more practical account of the influence of lateral waves, Fig. 9 shows the directivity obtained when all the source is composed of all the cut-on modes of the duct with uncorrelated phases and an equal energy distribution. This source is often used as a first approximation for broadband fan noise. When compared with the exact near-field solution, the far-field approximation appears to overpredict the amplitude of the peak by more than 1 dB and its location by more than 6°. Also the pressure amplitude inside the cone of silence is significantly underestimated. The use of correction terms to account for the fastest mode of the jet improves the results especially inside the cone of silence.

C. Directivity and rate of decay

Now that the lateral waves have been identified in the analytical solution, it is interesting to investigate their properties by inspecting Eqs. (9) and (10). Firstly, these equations can be simplified by assuming $Q \approx 1$. For Eq. (9) we can also

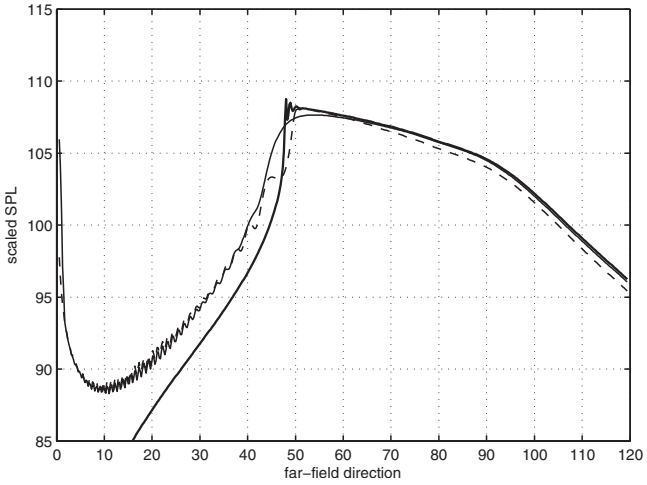


FIG. 9. Sound pressure level at $R=38$ for broadband noise: Far-field approximation (thick solid line); near-field solution (thin solid line); corrected far-field solution (dashed line).

use $C(z)\sim -z^{-2}/2$ as $z\to\infty$ (this result can be obtained by using items 7.1.3 and 7.1.23 of Ref. 25). Therefore, the correction term f_p behaves asymptotically as follows:

$$f_p(R,\varphi)=\frac{-i^{m+1/2}\sqrt{1-M_0^2}A_p e^{-i\omega RS(\varphi)}}{2\pi s_p^3 R^2\sqrt{2\lambda_0(u_p)\sin\varphi}(1-M_0^2\sin^2\varphi)^{3/4}},$$

when $R\to\infty$. It appears then that this correction term decays like R^{-2} since A_p and s_p are independent of R. This is similar to what Suzuki and Lele obtained for the three-dimensional planar shear layer.[18,19] If one notes that $s_p^2=-i-i\cos(\hat{\varphi}-\hat{\varphi}_p)$, where $\hat{\varphi}_p=\cos^{-1}(\tau_p)$, then it follows that the magnitude of s_p is minimum when $\hat{\varphi}=\hat{\varphi}_m=\mathrm{Re}(\hat{\varphi}_p)$. Due to the s_p^{-3} factor, this correction term has a strong directivity around $\hat{\varphi}_m$ and becomes weaker as the far-field direction moves away from $\hat{\varphi}_m$.

The other correction term g_p is zero for any angle greater than $\hat{\varphi}_c$ since the path of steepest descent does not cross the pole for these directions. The critical angle $\hat{\varphi}_c$ is the direction for which s_p is real, and the pole u_p lies on the path of steepest descent. The decay of g_p is given by the factor $e^{-\hat{\omega}\hat{R}s_p^2}/\sqrt{r}$ in Eq. (10). The far-field direction $\hat{\varphi}_m$ is again of interest since it also corresponds to s_p^2 being purely imaginary. So for $\hat{\varphi}=\hat{\varphi}_m$ the correction term would decay like $r^{-1/2}$. However, since $\hat{\varphi}_m$ is greater than $\hat{\varphi}_c$, this algebraic decay of g_p is not observed, and in fact g_p exponentially decays for any direction below $\hat{\varphi}_c$. Note, however, that for relatively small R the actual rate of the exponential decay can be very small, and thus the amplitude of the lateral waves can remain significant everywhere inside the cone of silence (see, for instance, the results in Fig. 5). However, as R increases, the rate of exponential decay increases and g_p remains significant only for directions close to $\hat{\varphi}_c$.

From a physical point of view, the lateral can be understood as conical waves (i.e., plane waves in the r-z plane) generated at the vortex sheet. The angle φ_c at which these plane waves are radiated is similar to the Mach angle of the disturbances inside the jet. Although these waves are present in the cone of silence in the near field, they will eventually leave the cone of silence when propagating to the far field and will generate a narrow beam of sound at the critical angle φ_c (this is the sharp peak observed in Fig. 5).

Finally, it should be noted that both f_p and g_p will eventually become negligible compared to the standard far-field approximation for any direction. Thus Eq. (7) represents the leading-order term in the far-field approximation.

D. Amplitude

The condition for the presence of lateral waves can also be inferred from Eqs. (9) and (10). The correction terms are proportional to the amplitude of the mode of the jet represented by the residue A_p of $P(u)$. Therefore, the fastest modes of the jet must be strongly excited by the source of sound for the lateral waves to be efficiently generated.

In the case of the Munt problem, Eq. (3) shows that A_p is proportional to $1/(u_p-\mu_{mn}^+)$. This confirms the observation made in Sec. IV A that lateral waves are efficiently generated by well cut-on modes since their axial wavenumbers are close to those of the fastest jet modes.

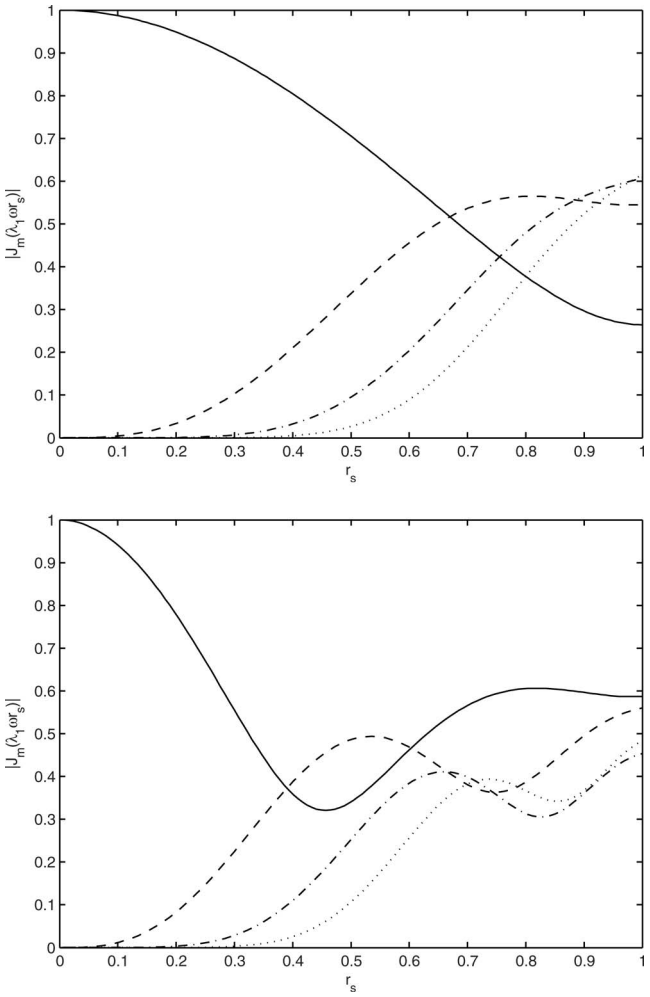


FIG. 10. Magnitude of $J_m(\lambda_1(u_p)\omega r_s)$ as a function of the point source position for different modes of the jet. Solid line: $m=0$; dashed line: $m=3$; dot-dashed line: $m=6$; dotted line: $m=9$. Top: fastest modes; bottom: second fastest modes.

For the plug flow jet, the location of the point source influences the magnitude of A_p through the factor $J_m(\lambda_1(u_p)\omega r_s)$. Its magnitude is plotted in Fig. 10. For a point source close to the axis, only the axisymmetric mode $m=0$ is excited, while higher-order modes are more efficiently generated when the source moves closer to the vortex sheet.

E. Discussion

It is generally assumed that the far-field asymptotic solution can be applied at several wavelengths from the duct exhaust. However the results in Fig. 5 show that inside the cone of silence the exact solution is still very different from the asymptotic solution at large distances from the jet exhaust. This is because a large portion of the vortex sheet acts as a source of sound for the ambient medium. Therefore, to define the far-field region where this approximation is valid, it is not appropriate to consider only the distance from the duct exhaust. It is preferable instead to think of the vortex sheet as a distributed source of sound that radiates in the ambient medium.

The presence of lateral waves can also have some consequences for computational aeroacoustics. Firstly, the sharp

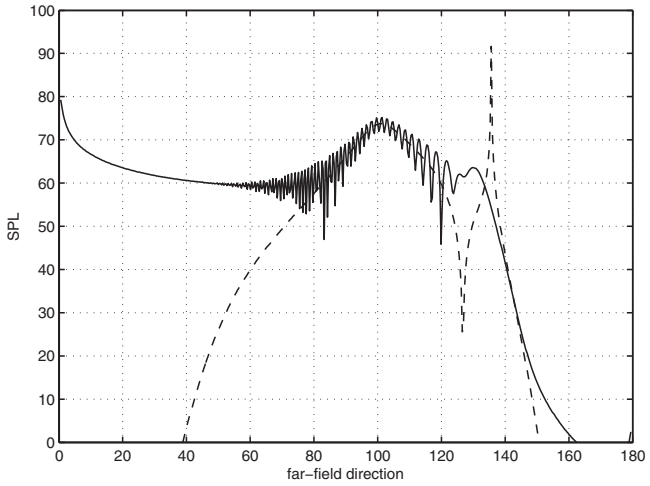


FIG. 11. Standard far-field approximation for a point source in a jet (dashed line). Modified approximation including the contribution from the channeled mode (solid line).

directivity peaks observed in Figs. 5 and 6 are the results of a precise interference between sound waves generated over a large portion of the vortex sheet. It is likely that even a small amount of error present in a numerical simulation will dissolve this interference and result in a smoother directivity pattern. Secondly, far-field solutions are generally obtained from numerical simulations by means of Kirchhoff or Ffowcs-Williams Hawkings surfaces. For problems similar to the Munt problem, an open cylindrical surface is generally used. The difficulty introduced by the lateral waves is that this cylinder has to extend far away from the duct exhaust to encompass most of the sound radiated from the jet. This implies that very large computational domains should be used.

V. CHANNELED JET MODES

Another feature that deserves a closer inspection is the presence of sharp peaks in the forward arc of the far-field approximation. An example is shown in Fig. 11 with the problem of a point source in a plug flow jet. The jet Mach number is 0.8 with unit density and sound speed ratios. The ambient medium is at rest and the Helmholtz number is 20. The source position is $r_s=0.9$ with an azimuthal order $m=19$. A peak significantly larger than the main lobe is observed at 136° in the analytical far-field solution. Note that similar peaks can also be observed in the Munt solution.

This feature originates from the presence of a zero of $L(u)$ close to the real axis between u_0^- and u_0^+. This zero represents a mode of the jet propagating in the positive axial direction with a slowly decaying amplitude since its axial wavenumber is almost real (the statement in Ref. 4 that such zeros lie exactly on the real axis was incorrect). For this mode, there is a very slow transfer of acoustic energy to the ambient flow, and the mode is "channeled" inside the jet. It is useful to have a closer look at the condition under which a mode of the jet is able to radiate sound in the ambient flow. To that end, consider relatively high frequencies and azimuthal orders so that the curvature of the jet is negligible on the acoustic length scale. The vortex sheet can then be thought of as a planar surface with some prescribed oscillations radiating in the semi-infinite ambient flow (this is the so-called wavy wall problem). The dispersion relation in the ambient flow can be written as

$$(\omega - k_z M_0)^2 = k_z^2 + k_r^2 + k_\theta^2,$$

where k_r is the wavenumber in the direction normal to the wavy wall and k_z and k_θ are wavenumbers in the directions parallel to the wall. For the jet mode we have $k_z=\omega u$ and $k_\theta=m/r=m$ since the jet radius is 1. For the mode to radiate sound k_r should be real, which yields the condition $\hat{u}_0^- < u < \hat{u}_0^+$ on the reduced axial wavenumber u with

$$\hat{u}_0^{\pm} = \frac{\pm\sqrt{1-(1-M_0^2)m^2/\omega^2} - M_0}{1-M_0^2}. \tag{13}$$

When this condition is satisfied, the mode radiates sound in the ambient flow, which is to say that the total phase speed is supersonic. Otherwise the mode induces only an evanescent disturbance in the vicinity of the vortex sheet but does not transfer energy to the ambient medium. In that case the jet effectively acts as a waveguide and the disturbances are channeled inside the jet.

With the far-field approximation, the saddle point moves from u_0^+ to u_0^- when φ varies from 0 to π. This range of values of u corresponds to disturbances with supersonic *axial* phase speed. The peak observed in Fig. 11 corresponds to a mode of the jet with a subsonic total phase speed (i.e., acoustic waves are channeled inside the jet and does not radiate sound) but with a supersonic axial phase speed (so it lies in the region spanned by the saddle point). Channeled modes can appear in the standard far-field solution when their axial wavenumbers are in the range $u_0^- < u < \hat{u}_0^-$. From Eq. (13) it appears that this is more likely to happen for large azimuthal orders m.

The physical principles associated with the channeled waves are the following. The mode of the jet involved propagates downstream but only very weakly radiates sound in the ambient medium due to its subsonic total phase speed. The key is that the axial wavenumber of this mode corresponds to disturbances propagating in the upstream direction in the ambient medium. So although the waves transmitted in the ambient medium are originally located downstream of the source, they will propagate in the upstream direction and they will eventually evolve into a narrow beam of sound observed in the forward arc of the standard far-field approximation. It should be noted that this situation is quite different from the lateral waves described in the previous section. Both lateral waves and channeled waves are due to disturbances propagating on long distances inside the jet (with a limited but nonzero transmission of sound through the vortex sheet). However, the rate of decay of channeled waves was found to be at least an order of magnitude smaller than that of lateral waves.

As explained in Sec. III B, the standard far-field approximation (8) is inaccurate when poles are in the vicinity of the path of steepest descent and when the contributions of poles crossed when deforming the contour are important. This is the case with channeled modes, and one has to in-

clude correction terms (9) and (10). As shown in Fig. 11, these correction terms remove the peak observed in the standard approximation. In addition a strong pressure field is added inside the cone of silence owing to the fact that in the near field these waves are generated downstream of the source. The behavior of the correction terms has already been discussed above, and the conclusions from Secs. IV B–IV D also apply in the case of channeled waves and are not repeated here.

VI. CONCLUDING REMARKS

The present analysis shows that the standard definition of the far field does not necessarily apply to problems involving sound propagation through a plug flow jet. The distance at which the standard far-field asymptotic solution becomes valid depends not only on the frequency and on the ambient flow properties but also on the properties of the jet and on the nature of the source of sound. For instance, lateral waves are generated by acoustic disturbances propagating over many wavelengths inside the jet with only limited decay in amplitude. As a result, some care should be exercised when far-field solutions are used for the plug flow jet and the Munt problem. If these analytical solutions are used as simple engineering models to assess trends between different configurations, the far-field approximation should provide reliable information (except maybe when the source of sound excite only well cut-on modes). If, however, these models are used to benchmark computational methods, it is recommended to use the exact near-field solutions as a comparison with the far-field asymptotic solutions might be inaccurate.

For an observer in the ambient flow, it can be misleading to understand the radiated field only in terms of the source of sound inside the jet (either a duct modes or a point source). Rather it is more appropriate to consider the corresponding distributed source at the vortex sheet. The extent of this distributed source is directly related to the amount of sound reflection inside the jet and to the nature of the source.

In practice the mixing layer of a jet has a finite thickness and grows axially. The extent to which the features observed with the plug flow jets translate to more realistic flow profiles remains to be investigated. It will be worth investigating the properties of lateral waves and of the far-field approximation in the context of round jets with finite mixing layers.

ACKNOWLEDGMENTS

This work was supported by the European Community through the TURNEX project (Technical officer: Daniel Chiron). The author is grateful to Dr. Brian Tester, Dr. Sjoerd Rienstra, and Dr. Chris Powles for several helpful discussions during the course of this work.

APPENDIX A: BRANCH CUTS OF THE INTEGRANDS

For the plug flow jet the integrand is given by Eq. (2). One can write

$$\frac{L(u)}{J_m(\lambda_1\omega r_s)} = \frac{J_m(\lambda_1\omega)}{J_m(\lambda_1\omega r_s)}\left[D_1(1-uM_1)^2\lambda_0 H_m'(\lambda_0\omega) - (1-uM_0)^2 H_m(\lambda_0\omega)\frac{\lambda_1 J_m'(\lambda_1\omega)}{J_m(\lambda_1\omega)}\right].$$

The Bessel function and its derivative can be written as infinite products (see item 9.5.10 of Ref. 25),

$$J_m(z) = \frac{z^m}{2^m m!}\prod_{n=1}^{\infty}\left(1-\frac{z^2}{j_{mn}^2}\right) \quad \text{for } m \geqslant 0,$$

$$J_m'(z) = \frac{z^{m-1}}{2^m (m-1)!}\prod_{n=1}^{\infty}\left(1-\frac{z^2}{j_{mn}'^2}\right) \quad \text{for } m > 0,$$

where j_{mn} and j_{mn}' are the zeros of $J_m(z)$ and $J_m'(z)$, respectively. It follows from these expressions that for any $m \geqslant 0$, both $J_m(\lambda_1\omega)/J_m(\lambda_1\omega r_s)$ and $\lambda_1 J_m'(\lambda_1\omega)/J_m(\lambda_1\omega)$ are regular functions of λ_1^2 alone and hence do not include the branch cuts of λ_1. It also follows that these two functions are nonsingular at $u_1^{\pm}$ where $\lambda_1 = 0$. Therefore, the branch cuts of λ_1 are not present in the integrand for the plug flow jet, which is also analytic at $u_1^{\pm}$.

For the Munt problem, the integrand is given by Eq. (3). The Wiener–Hopf kernel can be written as $K(u) = \lambda_0(u)\lambda_1(u)\hat{K}(u)$ with

$$\hat{K}(u) = D_1(1-uM_1)^2\frac{\lambda_1 J_m'(\lambda_1\omega)}{J_m(\lambda_1\omega)} - (1-uM_0)^2\frac{H_m(\lambda_0\omega)}{\lambda_0 H_m'(\lambda_0\omega)}.$$

From the analysis above, it appears that the branch cuts of λ_1 are not present in $\hat{K}$. The split functions can be written as $K_{\pm}(u) = \lambda_0^{\pm}(u)\lambda_1^{\pm}(u)\hat{K}_{\pm}(u)$, so the integrand now reads

$$P(u) = \frac{i\Psi_{mn}(1)\lambda_0^-(\mu_{mn}^+)\lambda_1^-(\mu_{mn}^+)K_-(\mu_{mn}^+)(1-uM_0)^2}{\omega(u-\mu_{mn}^+)\lambda_0(u)\hat{K}_+(u)H_m'(\lambda_0\omega)}.$$

It follows that the branch cuts of λ_1 do not appear in the integrand for the Munt problem.

APPENDIX B: APPROXIMATION OF THE ZEROS

We seek an approximation for the zeros of $L(u)$ that correspond to the fastest modes of the jet column. As shown in Fig. 2, it appears that these zeros are close to the axial wavenumbers μ_{mn}^+ of the duct modes given by $J'(\lambda_1\omega) = 0$. It is therefore convenient to rewrite $L(u) = 0$ as $f(u)J'(\lambda_1\omega) = 1$, where

$$f(u) = \frac{(1-uM_0)^2}{D_1(1-uM_1)^2}\frac{H_m(\lambda_0\omega)}{\lambda_0 H_m'(\lambda_0\omega)}\frac{\lambda_1}{J_m(\lambda_1\omega)}.$$

An approximation of a zero of $L(u)$ can be obtained by substituting u by $\mu_{mn}^+ + \delta u$ and by using a first-order Taylor expansion with respect to δu. This yields $\delta u = [\omega f(\mu_{mn}^+)\lambda_1'(\mu_{mn}^+)J_m''(j_{mn}')]^{-1}$. Equation (12) is obtained by using $\lambda_1\lambda_1' = -M_1C_1^2 - (1-M_1^2C_1^2)u$.

[1]J. D. Morgan, "The interaction of sound with a subsonic cylindrical vortex sheet," Proc. R. Soc. London, Ser. A **344**, 341–362 (1975).
[2]B. J. Tester and C. L. Morfey, "Developments in jet noise modelling-theoretical predictions and comparisons with measured data," J. Sound Vib. **46**, 79–103 (1976).
[3]R. M. Munt, "The interaction of sound with a subsonic jet issuing from a semi-infinite cylindrical pipe," J. Fluid Mech. **83**, 609–640 (1977).
[4]G. Gabard, B. J. Tester, and L. De Mercato, "Acoustic near- to far-field characteristics of nozzle radiation through plug flow jets," in 13th AIAA/CEAS Aeroacoustics Conference (2007), AIAA Paper No. 2007-3652.
[5]K. L. Gee, M. R. Sheperd, L. E. Falco, A. A. Atchley, L. S. Ukeiley, B. J. Jansen, and J. M. Seiner, "Identification of nonlinear and near-field effects in jet noise using nonlinearity indicators," in 13th AIAA/CEAS Aeroacoustics Conference (2007), AIAA Paper No. 2007-3653.
[6]K. Viswanathan, "Instrumentation considerations for accurate jet noise measurements," AIAA J. **44**, 1137–1149 (2006).
[7]M. S. Howe, "Transmission of an acoustic pulse through a plane vortex sheet," J. Fluid Mech. **43**, 353–367 (1970).
[8]J. W. Miles, "On the disturbed motion of a plane vortex sheet," J. Fluid Mech. **4**, 538–552 (1958).
[9]D. S. Jones, "Instability due to acoustic radiation striking a vortex sheet on a supersonic stream," *Proc. R. Soc. Edinburgh* **71**, 121–140 (1973).
[10]D. S. Jones, "The reflection of an acoustic pulse by a plane vortex sheet," Proc. Cambridge Philos. Soc. **74**, 347–362 (1973).
[11]D. S. Jones and J. D. Morgan, "The instability of a vortex sheet on a subsonic stream under acoustic radiation," Proc. Cambridge Philos. Soc. **72**, 465–488 (1972).
[12]D. G. Crighton and F. G. Leppington, "Radiation properties of the semi-infinite vortex sheet: The initial-value problem," J. Fluid Mech. **64**, 393–414 (1974).
[13]P. Gottlieb, "Sound source near a velocity discontinuity," J. Acoust. Soc. Am. **32**, 1117–1122 (1960).
[14]A. B. Friedland and A. D. Pierce, "Reflection of acoustic pulses from stable and instable interfaces between moving fluids," Phys. Fluids **12**, 1148–1159 (1969).
[15]M. E. Goldstein, "The low frequency sound from multipole sources in axisymmetric shear flows, with applications to jet noise," J. Fluid Mech. **70**, 595–604 (1975).
[16]M. E. Goldstein, *Aeroacoustics* (McGraw-Hill, New York, 1976).
[17]M. E. Goldstein, "High frequency sound emission from moving point multipole sources embedded in arbitrary transversely sheared mean flows," J. Sound Vib. **80**, 499–522 (1982).
[18]T. Suzuki and S. K. Lele, "Refracted arrival waves in a zone of silence from a finite thickness mixing layer," J. Acoust. Soc. Am. **111**, 716–728 (2002).
[19]T. Suzuki and S. K. Lele, "Green's functions for a source in a mixing layer: Direct waves, refracted arrival waves and instability waves," J. Fluid Mech. **477**, 89–128 (2003).
[20]G. Gabard and R. J. Astley, "Theoretical model for sound radiation from annular jet pipes: Far- and near-field solutions," J. Fluid Mech. **549**, 315–341 (2006).
[21]A. Demir and S. W. Rienstra, "Sound radiation from an annular duct with jet flow and a lined centerbody," in 12th AIAA/CEAS Aeroacoustics Conference (2006), AIAA Paper No. 2006-2718.
[22]B. Veitch and N. Peake, "Models for acoustic propagation through turbofan exhaust flows," in 13th AIAA/CEAS Aeroacoustics Conference (2007), AIAA Paper No. 2007-3543.
[23]S. W. Rienstra, "Acoustic radiation from a semi-infinite annular duct in a uniform subsonic mean flow," J. Sound Vib. **94**, 267–288 (1984).
[24]B. Noble, *Methods Based on the Wiener-Hopf Technique* (Pergamon, London, 1958).
[25]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964).
[26]L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (IEEE, New York, 1994).
[27]L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media. II. Point Sources and Bounded Beams* (Springer-Verlag, New York, 1992).
[28]L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1959).

Transitional wall pressure fluctuations on axisymmetric bodies

Chinsuk Hong,[a),b)] Ku-Kyun Shin,[c)] and Jae-Jin Jeon[d)]
Agency for Defense Development, P.O. Box 18, Chinhae, Kyungnam 645-600, Republic of Korea

Sang-Youn Kim[e)]
Yonam Institute of Digital Technology, 78 Gajoadong, Jinju, Kyungnam 660-300, Republic of Korea

(Received 5 November 2007; revised 22 August 2008; accepted 25 August 2008)

This paper reports the characteristics of wall pressure fluctuations in the transition region of the flow over axisymmetric bodies. Measurements are conducted in a low noise wind tunnel to obtain the boundary-layer flow field, the local static pressures, and the wall pressure fluctuations using hot wires, micromanometers and flush-mounted microphones. The spatial and the temporal developments of Tollmien–Schlichting (T-S) waves were first observed in the time domain. The wall pressure fluctuations in the transitional boundary-layer flows were of intermittent pulses and intensified as they were convected downstream at typically 63% of the upstream velocity. The Wigner–Ville distributions are then obtained to examine the energy evolution jointly in time and in frequency. The center frequency of the T-S wave is decreased with increases of the boundary-layer thickness and the bandwidth of the energy distribution is broadened as the local Reynolds number is increased. Finally the nondimensional spectra of the transitional wall pressure fluctuations scaled on the outer variables was obtained and it was then found that the characteristic frequency of the T-S wave was related to the outer variables as $\omega\delta^*/U_\infty \approx 0.2$. The peak level of the pressure fluctuations during late transition at the characteristic frequency is about 10 dB higher than that of the fully developed flow. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2988291]

PACS number(s): 43.28.Ra [GCL] Pages: 2767–2773

I. INTRODUCTION

The flow characteristics around a body have long been of fundamental interest in the study of fluid dynamics. Such investigations are critical in the design and development of practical objects such as military weapons and vehicles. When a body (in current study, an axisymmetric body) moves through a fluid, it usually forms a boundary layer on its surface due to the viscosity of the fluid. In addition, the boundary-layer flow changes from laminar to turbulent flow with a distinct transition, which could produce a fluid mechanical instability (e.g., pressure fluctuations) on the surface of the body. Basically, laminar flow is initially formed near the stagnation point on the nose of the axisymmetric body. No pressure fluctuation exists during laminar flow unless flow separation occurs. Then, the transition boundary-layer flow is produced and developed in the direction of the adverse pressure gradient, i.e., $\partial P/\partial x > 0$, where P is the static pressure and x is the downstream distance. Once small pressure disturbances are produced in the initial transitional boundary-layer flow, they may decay. However, farther downstream where the boundary-layer thickness is growing, wavelike disturbances having a characteristic frequency are produced due to transforming the external disturbances into internal instability oscillations of the boundary layer. These disturbances, taking the well-known form of *Tollmien–Schlichting wave* (T-S wave), may not decay. In relatively quiet flows, the initial amplitude of the T-S waves is insufficient to provoke immediate transition. T-S waves must first amplify in the boundary layer to trigger nonlinear effects, which is the main characteristic of the transition process. The flow eventually becomes a fully developed turbulent boundary-layer flow. The wall pressure fluctuations induced by the complicated movement of fluid particles in the boundary layer become excitation sources to the body. These sources hence produce structural vibrations generating noise outside and/or inside the body. This mechanism of noise generation is referred to as *transition-induced flow noise* (hereafter denoted by flow noise)—for more details, see Lauchle (1980a), Lauchle (1980b), Lauchle (1981), Lauchle (1977), and Arakeri *et al.* (1991).

The wall pressure fluctuations under a fully developed turbulent boundary-layer flow was first measured on a flat plate with no pressure gradient by Willmarth (1956). Since then, many researches on wall pressure fluctuations have been thoroughly conducted theoretically (Kraichnan, 1956) and experimentally (Blake, 1986a; Blake 1986b), but the results pertain mostly to flat plates. Flow over axisymmetric bodies is characterized by the behavior of the pressure gradient along the surface of the body. The pressure is initially favorable gradient, i.e, $\partial P/\partial x < 0$, around the nose of the axisymmetric body and then become adverse farther downstream. Here, P is the static pressure and x is the arc-length coordinate along the body surface. Spanwise curvature effects further complicate the boundary-layer development. The flow instability within the transition region of the boundary layer at the nose of an axisymmetric body has not re-

[a)]Present address: School of Mechanical Engineering, Pusan National University, Republic of Korea.
[b)]Author to whom correspondence should be addressed. Electronic mail: cshong@pusan.ac.kr
[c)]Electronic mail: kkshin@add.re.kr
[d)]Electronic mail: jjjeon@add.re.kr
[e)]Electronic mail: kkimsy@yc.ac.kr

ceived a great deal of attention in the literature (Lauchle, 1991). However, several research findings on turbulent boundary-layer wall pressure fluctuations over axisymmetric bodies have been presented. Perraud (1989) measured the wall pressure fluctuations on a long ellipsoidal nose and showed that the convection velocity of the wall pressure fluctuations in the transition region is about 45% of the upstream velocity. He also showed that the local wall pressure fluctuations in the transition region are possibly 10–25 dB higher than those of the turbulent boundary layer. Katz *et al.* (1990) and Abarbanel *et al.* (1994) conducted a buoyancy-propelled vehicle test (BVT) to measure wall pressure fluctuations on the surface of an axisymmetric body. They showed that the root-mean-square value of the wall pressure fluctuations in the transition region is about 18 dB higher than that in a fully developed turbulent flow. The strong intensity of wall pressure fluctuations in transitional boundary-layer flows is due to the rapid development of the T-S wave into a state of high nonlinearity (Katz *et al.*, 1990). However, there are other sources of wall pressure fluctuations in the transitional boundary layer. The level of wall pressure fluctuation in the transition region is further increased due to the variation of the angle of attack while the body is vertically running, as pointed out by Gal-el Hak (1987). He showed that a small variation of the angle of attack separates the boundary-layer flow and that the separation leads to an increase of the level of wall pressure fluctuations. The separation and reattachment of the boundary-layer flow are essential features of the unsteady flow at the nose of an axisymmetric body. Arakeri (1975) experimentally identified wall pressure fluctuations induced by either transition or separation on an axisymmetric nose and showed that the pressure fluctuations in the reattachment region are ten times higher than those in the fully developed turbulent boundary-layer flow.

This paper investigates the transitional wall pressure fluctuations on an axisymmetric body over which the boundary layer develops naturally from laminar to turbulent flow, including a distinct transition region without separation. Understanding of the transitional wall pressure fluctuations is essential for analyzing the mechanism of flow noise generation. The wall pressure fluctuations produced by transitional flows are basically unsteady and inhomogeneous, whereas those of fully developed turbulent flows are steady and homogeneous. It is also important to find a relationship between the wall pressure spectrum and the basic flow conditions such as the upstream inlet velocity, the boundary-layer thickness, and the displacement thickness. In this study, the measurement of the wall pressure fluctuations is conducted under the condition of a stable zero angle of attack and upstream velocity for prevention of separation. The measured transitional wall pressure fluctuation signal is characterized by the temporal feature of the signal, the frequency spectra, and the time-frequency spectra. Finally, the relationship between the transitional wall pressure fluctuation and flow conditions is obtained by using dimensionless scaling techniques (Blake, 1970; Keith *et al.*, 1991; Lauchle, 1980b).

In Sec. II, the experimental setup such as the test facility, test model, and measurement instrumentation is explained.

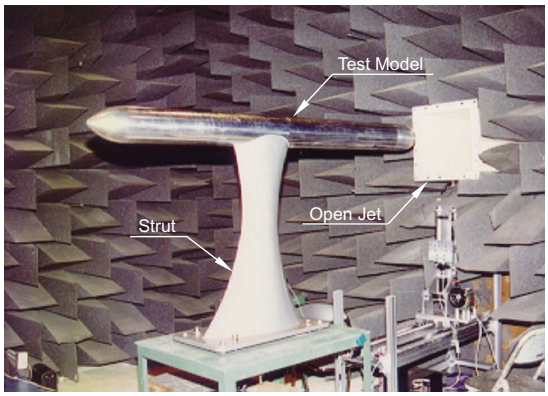


FIG. 1. (Color online) Axisymmetric body in the anechoic chamber of low noise wind tunnel having 400×400 mm^2 open jet. The body is supported by a strut having NACA0018 section.

The characteristics of the transitional wall pressure fluctuations are discussed in Sec. III in terms of the temporal signal, the frequency spectra, and the time-frequency spectra. Conclusions from this study are given in Sec. IV.

II. EXPERIMENTAL SETUP

A. Test facility and test model

The measurement of wall pressure fluctuations was conducted in a low noise wind tunnel,[1] as shown in Fig. 1. The wind tunnel is an open-jet type having the dimension of 400×400 mm^2. The maximum flow speed is 56 m/s. The turbulent intensity is less than 1% (Kim *et al.*, 1998).

Figure 2 shows the schematic shape of the model, which consists of a blunt nose, a cylindrical shell, and a sharp tail. The model was made of aluminum. The dimensions of the model were 2 m long, 160 mm in diameter, and 10 mm in thickness. The shape of nose was formed as a flat surface in the front followed by an elliptical curve that is tangent to the front flat on the major axis and tangent to the cylindrical body on the minor axis. To install the model in the anechoic chamber, a strut whose section was the shape of NACA0018 was used, so that the flow disturbance due to the strut could be minimized.

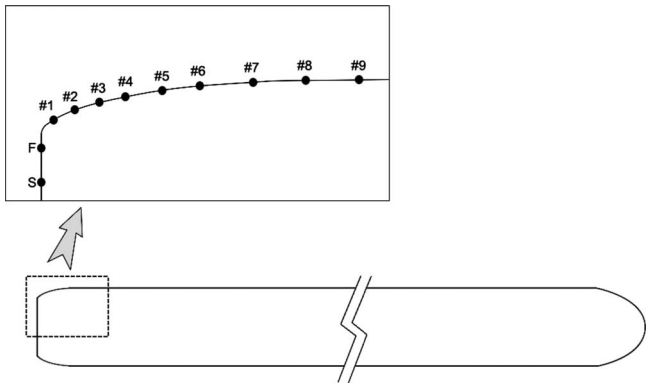


FIG. 2. Shape of the test model and locations of the sensors: S denotes the stagnation point at which a pressure tap is installed and F denotes the front face of the test model at which a pressure tap is also installed. From 1 to 9, pressure taps and microphones are installed on different models with the same shape and dimensions.

TABLE I. Sensor positions measured with the arc length along the surface with respect to the stagnation point.

		$Re_s/10^5$			δ^* (m)		
Sensors	Arc lengths (m)	**20 m/s**	**30 m/s**	**40 m/s**	**20 m/s**	**30 m/s**	**40 m/s**
S	0	0	0	0	0.0061	0.001	0
F	0.0497	0.569	0.854	1.1389	0.0003	0.0002	0.0002
1	0.0612	1.1065	1.6625	2.2196	0.0008	0.0007	0.0007
2	0.0764	1.2949	1.9358	2.5842	0.0017	0.0015	0.0015
3	0.0845	1.4051	2.0979	2.798	0.002	0.0017	0.0017
4	0.0978	1.5814	2.3748	3.1658	0.0023	0.0023	0.0023
5	0.1142	1.8185	2.7238	3.6293	0.0026	0.0026	0.0026
6	0.1312	2.0464	3.0743	4.1031	0.003	0.003	0.003
7	0.1568	2.4171	3.6273	4.8376	0.0035	0.0035	0.0035
8	0.1828	2.7734	4.1609	5.5489	0.004	0.004	0.004
9	0.2063	3.0466	4.5692	6.0926	0.0053	0.0046	0.0046

B. Instrumentation

Hot wires, micromanometers, and microphones are used to measure the boundary-layer flow field, the local static pressures, and the wall pressure fluctuations. The velocity profile is measured using the hot wires (TSI hot-wire probe and wire TSI 1210 T-1.5) whose location is precisely controlled by the traversing system shown in Fig. 1. This velocity profile measurement is performed at every location, as marked in Fig. 2. The micromanometers and microphones were mounted in the form of an array along the surface of the test model. The micromanometers were used to measure the static pressures on the surface along the body from the stagnation point to cylindrical part, where the diameter almost reaches the maximum. The locations of the sensors are listed in Table I and shown in Fig. 2. The locations are represented in terms of the arc length from the forward stagnation point. Nine pressure taps (1–9) were installed along the ellipsoidal surface, as shown in Fig. 2. Two more pressure taps ("S" and "F") were also installed on the flat surface at $x=0$: one at the stagnation point and the other at $(x,r)=(0,49.6\ \text{mm})$. The diameter of the pressure taps was 1 mm. The static pressures were read by a micromanometer of Furness Co. Ltd (FCO12 Model 3). The micromanometer can measure the static pressure over 2 ranges of ± 19.99 mm H_2O and ± 199.9 mm H_2O with the accuracy of $\pm 0.5\%$ FSD (± 1 digit).

Microphones were also used to measure the wall pressure fluctuation on the surface of the ellipsoidal part of the axisymmetric body. Eleven 1/8 in. condenser microphones were flush mounted at the locations (S, F, and 1–9 in Fig. 2), where the flow develops from laminar to turbulent with a distinct transition region at the flow speeds of interest. Note that the locations of the microphones are the same as those of the micromanometers and of the hot-wire measurements. The measurement was affected by the size of the diaphragm of the microphones. Since the size of the diaphragm of the microphones is finite, the signal captured by the diaphragm is spatially averaged over the area of the diaphragm (Lueptow, 1993). At low frequencies, the averaged signal can be equivalent to the signal at a point. The maximum frequency of the measurement using the flush-mounted microphone can be estimated. We can obtain the spatially averaged signal without any loss for the signal whose half wavelength is greater than the diameter of the diaphragm. Assuming that a spot travels with the speed of $0.5U_\infty$, hence, the maximum frequency $f_{\max}$ will be 1600, 3400, and 6400 Hz for 20, 30, and 40 m/s, respectively. The maximum frequencies estimated for the speeds of the experiment are enough to cover the frequency range of interest in this study. At higher frequencies than the maximum frequencies, the measured signal is unreliable due to the spatial averaging. The pin-hole system of Lueptow (1993) and Blake (1970) is not used in this study to obtain further reduction of the sensing area. The signals from the microphones were conditioned by using preamplifiers. Using low pass filters (Krone-Hite, KH-3100A), the low frequency contaminations due to machinery for tunnel operations were removed. A multichannel spectrum analyzer (B&K PULSE System) was used to clarify the characteristics of the wall pressure fluctuations. The measured signals were digitalized and stored at a personal computer. Further data processing was conducted using mainly the signal processing toolbox of MATLAB®.

III. RESULTS AND DISCUSSION

A. Static pressures and flow regime

The static pressures at the specified locations were measured at the freestream velocities of 10, 20 30, 40, and 50 m/s and the static pressure coefficients were then calculated as

$$C_p = \frac{P - P_\infty}{\frac{1}{2}\rho U_\infty^2}, \tag{1}$$

where P is the measured static pressure, P_∞ is the reference pressure, ρ is the density of air, and U_∞ is the freestream velocity.

Figure 3 shows the variation of the static pressure coefficients with locations along the body surface. The pressure coefficient decreases downstream as the arc length increases from the pressure tap S to 1, showing the favorable pressure gradient. This favorable pressure gradient changes into the adverse pressure gradient downstream of static pressure tap 1. This means that the fluid particles accelerate to this point and then decelerate afterward. It is noted that the pressure

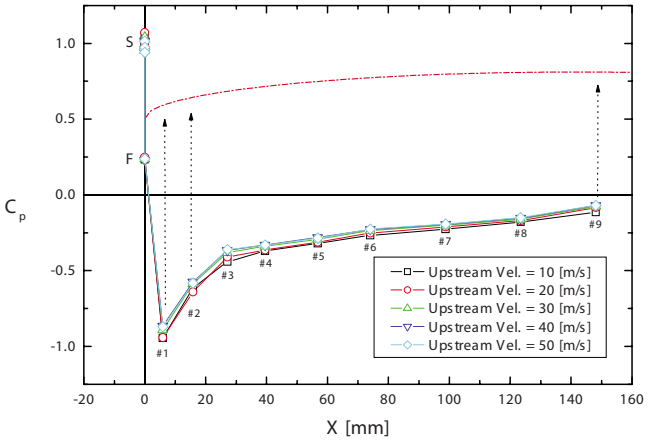


FIG. 3. (Color online) Variation of static pressure coefficients for upstream velocities of 10, 20, 30, 40, and 50 m/s. S denotes the stagnation point and F denotes the front face. The dot-dashed line shows the shape of the axisymmetric body.

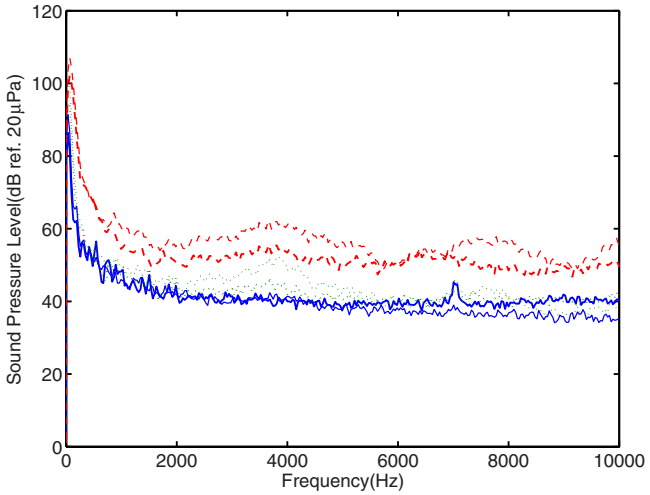


FIG. 4. (Color online) Background noise levels measured at the microphones S (thin lines) and F (thick lines) with the whole test body supported by the strut for open-jet speeds of 20 m/s (solid lines), 30 m/s (dotted lines), and 40 m/s (dashed lines).

distribution shows a weak dependence on the freestream velocity. The speed dependence might be caused by horizontal buoyancy, as described by Lauchle (1979).

The velocity profiles at the sensor locations (S, F, and 1–9) are measured using a hot-wire probe for three upstream velocities of 20, 30, and 40 m/s. They are then used to calculate the displacement thicknesses by

$$\delta^* = \int_0^\delta \left(1 - \frac{u}{U_\infty}\right) d\eta, \tag{2}$$

where δ is the boundary-layer thickness and u is the velocity profile, at each sensor location. The coordinate η is normal to the surface. The displacement thickness and local Reynolds number at each sensor location are listed in Table I for the three upstream velocities. The local Reynolds number is defined by

$$\mathrm{Re}_s = \frac{U_\infty s}{\nu}, \tag{3}$$

where s is the arc length measured from the stagnation point and ν is the kinematic viscosity. The critical local Reynolds number is about 3×10^5 for blunt-nosed bodies (White, 2003). The local Reynolds numbers listed in Table I for sensor locations can be compared with the critical Reynolds number to estimate the flow regime. It can be surmised from this comparison that the sensors are located under the boundary layer undergoing transition.

B. Background noise levels

Before measuring the wall pressure fluctuations using the flush-mounted microphones, we measured the background noise with installation of the body in the test section at flow speeds of 20, 30, and 40 m/s. This measurement is to ensure that the noise generated by, for example, tunnel operating machinery, the strut, and the tail of the body, is not dominating the measured wall pressure fluctuations. The microphones at S and F are used to measure the background noise since they are located under the laminar boundary layer so that no pressure fluctuations exist. Note that the laminar flow at these two locations can be confirmed by the comparison of the local Reynolds number listed in Table I to the critical Reynolds number of 3×10^5. Figure 4 shows the measured background noise levels at the microphones S (thin lines) and F (thick lines) with the whole test body supported by the strut for open-jet speeds of 20 m/s (solid lines), 30 m/s (dotted lines), and 40 m/s (dashed lines). It can be seen that the background noise level is about 105 dB *re* 20 μPa at about 100 Hz. This high background noise is mostly contaminated by tunnel operational machinery noise. At the high frequencies above 1 kHz, the background noise is about 60 dB *re* 20 μPa at 40 m/s, and varies with the upstream velocity. This noise can be affected by the noise generated by the interaction of flow with the open-jet, the strut, and the tail of the body. The measurement of wall pressure fluctuations is restricted by the level of this background noise.

C. Wall pressure fluctuations

The wall pressure fluctuations at the microphone positions 4, 6, 7 and 8 were analyzed for the upstream velocities of 20, 30, and 40 m/s where the boundary-layer flow naturally developed from laminar to turbulent including a distinct transition region in the range of the microphone locations. To show a clear development process of the wall pressure fluctuations, we examined the time history of the wall pressure fluctuations in 0.1 s records with increases in the arc-length Reynolds number. Figure 5 shows the development of the wall pressure fluctuations with the arc-length Reynolds numbers. Note that we selected nine sets of data from the data for 12 Reynolds numbers given in Table I. We omitted data (wall pressure of laminar flows at 30 and 40 m/s) that are found to be contaminated by high background noise (as shown in Fig. 4) for the corresponding upstream velocities. It can be seen from Fig. 5 that the T-S wave is generated between $\mathrm{Re}_s = 1.16\times10^5$ and $\mathrm{Re}_s = 2.05\times10^5$ and then is amplified as fluid flows downstream. This amplification is due to the adverse pressure gradient as mentioned by Goldstein and Hultgren (1987). The flow at $\mathrm{Re}_s = 2.05\times10^5$ (microphone posi-

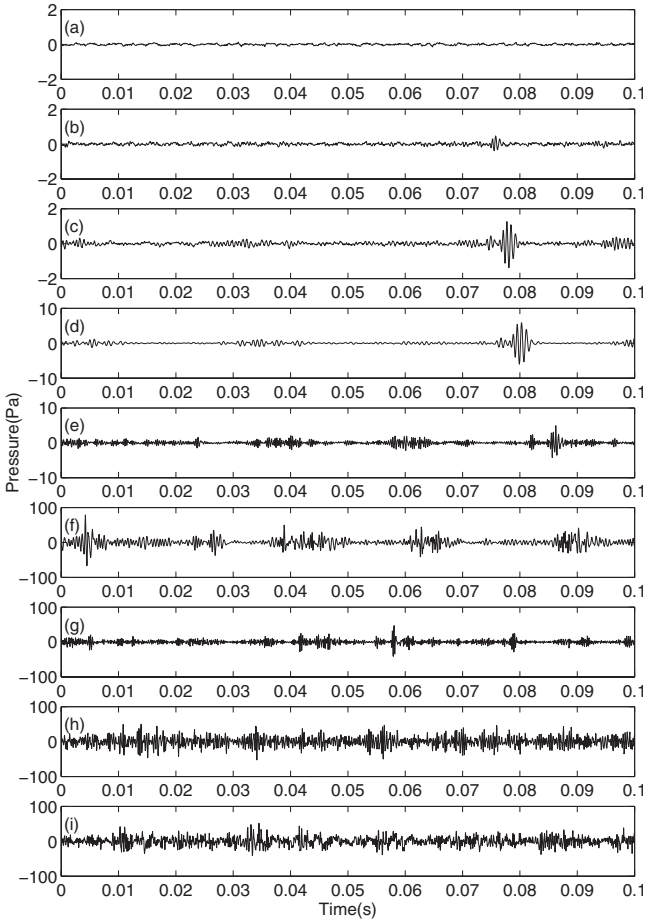


FIG. 5. Development of wall pressure fluctuations with the arc-length Reynolds numbers (a) $Re_s=1.11\times10^5$, (b) $Re_s=2.05\times10^5$, (c) $Re_s=2.42\times10^5$, (d) $Re_s=2.78\times10^5$, (e) $Re_s=3.07\times10^5$, (f) $Re_s=3.63\times10^5$, (g) $Re_s=4.10\times10^5$, (h) $Re_s=4.84\times10^5$, and (i) $Re_s=5.55\times10^5$.

tion 6 at $U_\infty=20$ in this case) is, therefore, in the initial transition and then develops to the late transition, downstream. The convection velocity is estimated from the time difference (2 ms in this case) between the pulses and the distance (26 mm in this case) between sensors and is 63% of the upstream velocity. This estimated average convection velocity is comparable to 30%–50% of the upstream velocity from the measurement conducted by Perraud (1989) on an axisymmetric body. It is noted that the estimated convection velocity is the average convection velocity because the spots grow in both the longitudinal and the transverse directions so that their leading and trailing edges travel at different speeds. As the Reynolds number was increased to $Re_s=3.63\times10^5$, the wall pressure fluctuations suddenly increased with higher frequency components. This can be considered as a feature of the late transition boundary-layer flow, as will be discussed more in Sec. III D. With the further increase of the Reynolds number to $Re_s=5.55\times10^5$, the flow shows eventually the feature of the turbulent boundary layer containing higher frequency components and more randomness. It can be seen that a temporal peaks at $Re_s=3.63\times10^5$ are higher than those at $Re_s=5.55\times10^5$.

D. Time-frequency spectrum

Interesting temporal features of the wall pressure fluctuations in the transition boundary layer are described above.

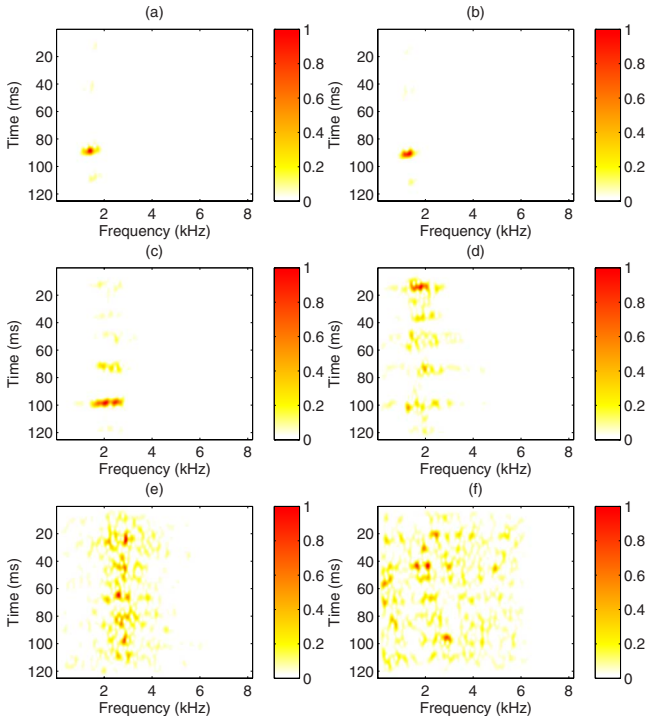


FIG. 6. (Color online) WVDs of wall pressure fluctuations at the measuring points 6, 7, and 8 for $U_\infty=20$, 30, and 40 m/s. (a) $Re_s=2.42\times10^5$, (b) $Re_s=2.77\times10^5$, (c) $Re_s=3.63\times10^5$, (d) $Re_s=4.16\times10^5$, (e) $Re_s=4.84\times10^5$, and (f) $Re_s=5.55\times10^5$.

These features suggest a time-frequency representation which shows the evolution of energy in time and in frequency simultaneously. This can be done effectively using the Wigner–Ville distribution (WVD) method (Quian and Shen, 1999; Boashash and Black, 1987), defined as

$$W(t,\omega)=\int_{-\infty}^{\infty}x\left(t+\frac{\tau}{2}\right)x^*\left(t-\frac{\tau}{2}\right)\exp(-2\pi\omega\tau)d\tau, \qquad (4)$$

where $W(t,\omega)$ is the WVD or the time-frequency spectrum, and $x(t)$ and $x^*(t)$ denote a signal and its conjugate, respectively. The WVD is a basic tool for joint time-frequency analysis without loss in resolution in either coordinate (time and frequency). There are simpler methods for analyzing signals in the time-frequency domain, for example, the short time Fourier transform. However, they fail to exhibit the same resolution capabilities in time and frequency. The short time Fourier transform uses a time windowed signal. The frequency resolution is inversely proportional to the window duration. The narrower the window, the better is the time resolution, but the worse is the frequency resolution and vise versa. The main advantage of WVD, hence, is that it does not suffer from this shortcoming.

We obtained WVD of the wall pressure fluctuations at 7 and 8 for $U_\infty=20$, 30, and 40 m/s. Figure 6 shows the WVDs of wall pressure fluctuations at the measuring points 6–8. Several features can be found from this time-frequency analysis. A very small pulse in the time history of the wall pressure fluctuation at $Re_s=2.47\times10^5$ shown in Fig. 5(a) gives a distinct energy distribution in the corresponding WVD. Comparing Figs. 5(a) and 5(b) obtained from the signals at 7 and 8 for $U_\infty=20$, respectively, it can be found that

the center frequency of the pulse is decreased as the fluid flows downstream in such a way that $f_c=1.3$ kHz for 7 and $f_c=1.5$ kHz for 8. This seems to be because the maximum eddy size is increased together with the thickness of the boundary layer as the fluid flows downstream. From the time difference of the peaks at the two stations, the average convection velocity is estimated to be 63% of the upstream velocity. As the arc-length Reynolds number increases, the spots occur more frequently. Finally, it can be pointed out from this analysis that the energy distribution in the frequency domain is broadened as Re_s is increased.

E. Scaling of wall pressure fluctuation spectra

In general, three scaling schemes are used to characterize the wall pressure fluctuations in turbulent boundary-layer flows. These schemes use the outer variables, inner variables, and mixed variables, respectively. The nondimensional spectrum is represented by

$$\Phi_p(\omega)/\rho^2\delta^*U_\infty^3 \text{ versus } \omega\delta^*/U_\infty \tag{5}$$

for the outer variable scaling,

$$\Phi_p(\omega)U_\tau^2/\tau_w^2\nu \text{ versus } \omega\nu/U_\tau^2 \tag{6}$$

for the inner variable scaling, and

$$\Phi_p(\omega)U_\infty/\tau_w^2\delta^* \text{ versus } \omega\delta^*/U_\infty \tag{7}$$

for the mixed variable scaling, where $\Phi_p(\omega)$ is the spectrum of the wall pressure fluctuation, τ_w is the wall shear stress, and $U_\tau=\sqrt{\tau_w/\rho}$ is the friction velocity.

According to Keith *et al.* (1991), high-frequency wall pressure fluctuations in turbulent boundary-layer flows can be well scaled using the inner-flow variables. On the other hand, low- to midfrequency wall pressure fluctuations of the turbulent boundary layer are better scaled using the outer-flow variables. It should be noted that these scaling schemes are mostly used for fully developed turbulent boundary-layer flows. Scaling for transitional wall pressure fluctuations are not readily found in the literature. However, since the main features of the transitional wall pressure spectrum shown in Fig. 6 occur at relatively low frequencies, the flow can be considered large scale, and the outer-flow scaling may apply.

Figure 7 shows the nondimensional spectra of the wall pressure fluctuations for a range of Reynolds numbers. These nondimensional spectra are compared with the result of Bull and Thomas (1976) for the scaled wall pressure fluctuations under fully developed turbulent boundary-layer flows. It is known that the nondimensional spectrum when $\mathrm{Re}_s=4.84\times10^5$ shows the typical spectral feature of fully developed turbulent boundary-layer flows. The deviation at higher frequencies ($\omega\delta^*/U_\infty<0.7$) is due to the spatial averaging over the diaphragm of the microphones as mentioned in Sec. II B. We can find a wide maximum centered at about $\omega\delta^*/U_\infty=0.2$. This hump in the frequency spectrum is generated by two dimensional T-S waves. The T-S waves in the initial transitional boundary-layer flow are slowly amplified as they convect downstream, as seen by the temporal features in Fig. 5 as well as in Figs. 7(a) and 7(b). Walker and Gostelow (1990) reported that two dimensional T-S waves exhibit high-frequency harmonics. In this measurement, however, such harmonics have not been observed.

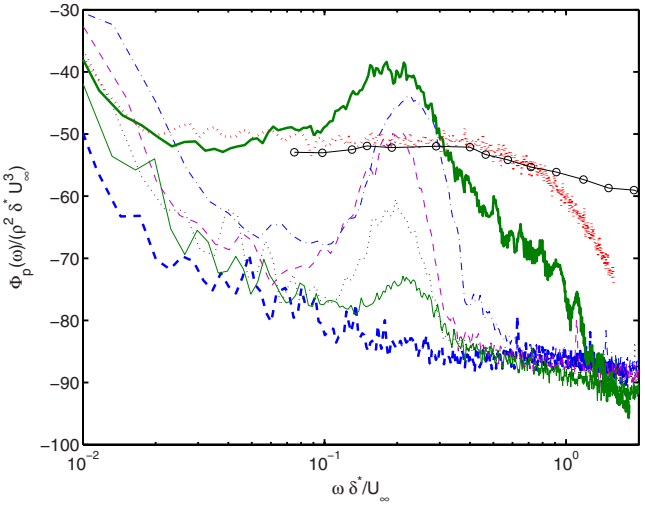


FIG. 7. (Color online) Nondimensional spectra of wall pressure fluctuations scaled on the outer variables when the arc-length Reynolds numbers are $\mathrm{Re}_s=1.11\times10^5$ (thick dashed), $\mathrm{Re}_s=2.05\times10^5$ (faint solid), $\mathrm{Re}_s=2.42\times10^5$ (faint dotted), $\mathrm{Re}_s=2.78\times10^5$ (faint dashed), $\mathrm{Re}_s=3.07\times10^5$ (faint dot-dashed), $\mathrm{Re}_s=3.63\times10^5$ (thick solid), $\mathrm{Re}_s=4.84\times10^5$ (thick dotted), and Bull and Thomas (1976) (circled).

The spectral hump due to the T-S waves is steadily intensified as Re_s is increased from the initial transition when $\mathrm{Re}_s=2.05\times10^5$ to the late transition when $\mathrm{Re}_s=3.63\times10^5$. The peak level of the wall pressure fluctuation at $\omega\delta^*/U_\infty=0.2$ when $\mathrm{Re}_s=3.63\times10^5$ is about 10 dB higher than the results when $\mathrm{Re}_s=4.84\times10^5$ where the flow is fully developed. A feature of the fully developed turbulent boundary-layer flow is the typical flat spectrum. Note that the nondimensional spectrum obtained in this study agrees well with the result of Bull and Thomas (1976) in the frequency range $\omega\delta^*/U_\infty\leq1$. At the low frequencies where $\omega\delta^*/U_\infty<0.1$, when $\mathrm{Re}_s=3.63\times10^5$, the spectrum level of the late transition is increased. Furthermore, at the high frequencies where $\omega\delta^*/U_\infty<0.3$, when $\mathrm{Re}_s=3.63\times10^5$, the level is also increased. This widening in bandwidth is due to the intensifying turbulent spots.

IV. CONCLUSION

This paper has described an investigation of the wall pressure fluctuations that occur under a transitional boundary layer formed on the nose of an axisymmetric body. The wall pressure fluctuations were measured on this body operating in a low noise wind tunnel using flush-mounted microphones. The spatial and temporal developments of T-S waves were first observed in the time domain. Further investigation of the T-S waves were then conducted by using the WVD method. The spectra of the transitional wall pressure fluctuations were found to scale on the outer-flow variables.

The main conclusions of this study are as follows.

(1) The wall pressure fluctuations in transitional boundary-layer flows were of intermittent pulses which are convected downstream at 63% of the upstream velocity.

(2) The intermittent wall pressure fluctuations in the transitional boundary-layer flows showed a broad hump in the frequency spectrum. Its center frequency was observed to decrease at downstream locations or as the thickness of the boundary layer increases.
(3) The center frequency of the broad peak in the transitional wall pressure fluctuation spectrum was found to scale well with outer-flow variables, i.e., $\omega\delta^*/U_\infty \approx 0.2$. This implies that the center frequency of the T-S waves is closely related to the thickness of the boundary layer and the upstream velocity.
(4) The level of the late transitional wall pressure fluctuations was about 10 dB higher than that of fully developed turbulent wall pressure fluctuations.

ACKNOWLEDGMENTS

Professor Gerald C. Lauchle of Pennsylvania State University reviewed this article and provided invaluable comments. His support is greatly appreciated.

[1]The low noise wind tunnel belongs to Korea Institute of Machinery and Materials (KIMM).

Abarbanel, H. D., Katz, R. A., and Frison, T. W. (**1994**). "Nonlinear analysis of high-Reynolds-number over a buoyant axisymmetric body," Phys. Rev. E **49**(5), 4003–4018.
Arakeri, V. H. (**1975**). "A note on the transition observations on an axisymmetric body and some related fluctuating wall pressure measurement," J. Fluids Eng. **97**, 82–87.
Arakeri, V. H., Satyanarayana, S. G., Mani, K., and Sharma, S. D. (**1991**). "Studies on scaling of flow noise received at the stagnation point of axisymmetric body," J. Sound Vib. **146**, 449–462.
Boashash, B., and Black, P. J. (**1987**). "An efficient real-time implementation of the Wigner-Ville distribution," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1611–1618.
Blake, W. K. (**1970**). "Turbulent boundary-layer wall-pressure fluctuations on smooth and rough walls," J. Fluid Mech. **44**, 637–660.
Blake, W. K. (**1986a**). *Mechanics of Flow-Induced Sound and Vibration* (Academic, New York), Vol. I.
Blake, W. K. (**1986b**). *Mechanics of Flow-Induced Sound and Vibration* (Academic, New York), Vol. II.
Bull, M. K., and Thomas, S. W. (**1976**). "High frequency wall pressure fluctuations in turbulent boundary layers," Phys. Fluids **19**, 597–599.
Gal-el Hak, M. (**1987**). "Unsteady separation on lifting surfaces," Appl. Mech. Rev. **40**, 441.
Goldstein, M. E., and Hultgren, L. S. (**1987**). "A note on the generation of Tollmien–Schlichting waves by sudden surface-curvature," J. Fluid Mech. **181**, 519–525.
Katz, R. A., Galib, T. A., and Cembrola, J. M. (**1990**). "Classical and nonlinear analysis of transition and turbulent boundary layer flow," in Proceedings of the Second International Symposium on Performance Enhancement for Marine Applications, pp. 119–127.
Keith, W. L., Hurdis, D. A., and Abraham, B. M. (**1991**). "A comparison of turbulent boundary layer wall-pressure spectra," ASME NCA, Vol. **11**, pp. 1–9.
Kim, J. S. *et al.* (**1998**). "Measurement of flow noise on an axisymmetric body using a low noise wind tunnel," Agency for Defense Development Technical Report No. NWSD-417-980039.
Kraichnan, R. H. (**1956**). "Pressure fluctuations in turbulent flow over a flat plate," J. Acoust. Soc. Am. **28**, 378–390.
Lauchle, G. C. (**1977**). "Noise generated by axisymmetric turbulent boundary layer flow," J. Acoust. Soc. Am. **61**, 694–703.
Lauchle, G. C. (**1979**). "Horizontal buoyancy effects on the pressure distribution of a body in a duct," J. Hydronaut. **13**, 61–67.
Lauchle, G. C., Eisenhuth, J. J., and Gurney, G. B. (**1980a**). "Boundary-layer transition on a body of revolution," J. Hydronaut. **14**, 117–121.
Lauchle, G. C. (**1980b**). "On the radiated noise due to boundary layer transition," J. Acoust. Soc. Am. **67**, 158–167.
Lauchle, G. C. (**1981**). "Transition Noise: The role of fluctuating displacement thickness," J. Acoust. Soc. Am. **69**, 665–671.
Lauchle, G. C. (**1991**). "Hydroacoustics of transitional boundary-layer flow," Appl. Mech. Rev. **44**, 517–531.
Lueptow, R. M. (**1993**). "Wall pressure transducer spatial resolution," NCA 15/FEDVol **168**, 49–55.
Perraud, J. C. (**1989**). "Studies of laminar-turbulent in air and water wall pressure fluctuations and acoustic emission from the turbulent intermittency," NCA **5**, 17–24.
Qian, S., and Shen, D. (**1999**). "Understanding the nature of signals whose power spectra change with time. Joint time-frequency analysis," IEEE Signal Process. Mag. **16**, 52–67.
Walker, G. J., and Gostelow, J. P. (**1990**). "Effect of adverse pressure gradients on the nature and length of boundary layer transition," ASME J. Turbomach. **112**, 198–205.
White, F. M. (**2003**). *Fluid Mechanics*, 5th ed. (McGraw-Hill, New York).
Willmarth, W. W. (**1956**). "Wall pressure fluctuations in a turbulent boundary layer," J. Acoust. Soc. Am. **28**, 1048–1053.

A new equation for the accurate calculation of sound speed in all oceans

Claude C. Leroy
IMM Vivaldi, 30 Chemin de la Baou, 83110 Sanary sur Mer, France

Stephen P. Robinson and Mike J. Goldsmith
National Physical Laboratory, Teddington, Middlesex TX11 OLW, United Kingdom

(Received 16 July 2007; revised 8 June 2008; accepted 26 August 2008)

A new equation is proposed for the calculation of sound speed in seawater as a function of temperature, salinity, depth, and latitude in all oceans and open seas, including the Baltic and the Black Sea. The proposed equation agrees to better than ± 0.2 m/s with two reference complex equations, each fitting the best available data corresponding to existing waters of different salinities. The only exceptions are isolated hot brine spots that may be found at the bottom of some seas. The equation is of polynomial form, with 14 terms and coefficients of between one and three significant figures. This is a substantial reduction in complexity compared to the more complex equations using pressure that need to be calculated according to depth and location. The equation uses the 1990 universal temperature scale (an elementary transformation is given for data based on the 1968 temperature scale). It is hoped that the equation will be useful to those who need to calculate sound speed in applications of marine acoustics.
[DOI: 10.1121/1.2988296]

PACS number(s): 43.30.Es, 43.30.Pc [RAS] Pages: 2774–2782

I. INTRODUCTION

The accurate calculation of the speed of sound in seawater is needed in various applications of underwater acoustics. Quite a number of equations for this purpose have been developed in the second half of the past century. Comprehensive reviews of the available data and equations have been published by Dushaw *et al.*[1] and Leroy.[2] The equations for sound speed may be clearly divided into two categories. The first are based on precise laboratory measurements where seawater samples of various salinities have been placed under a number of temperature and pressure combinations. These equations are complex polynomials best fits to the whole set of corresponding measured values. They use pressure (the measured parameter) as a variable, while many applications require the use of depth in the sea instead. In that case, a preliminary conversion of depth into pressure is required, with the conversion depending on location on the earth. The exact procedure, developed by oceanographers who needed extreme accuracy, is somewhat tedious and hard to apply. For this reason, a number of simpler equations were developed, most of the time by fitting the first ones over a limited domain of the variables. These simpler equations used depth Z (and eventually latitude Φ) instead of pressure P. The equations of this second category use a reduced number of terms and simpler coefficients, mainly because personal computers were rare or even not existing at the time they were developed.

This set of available equations is, however, not satisfactory to many users. The simple equations do not cover all existing waters and are not accurate enough for many applications. There are also problems when using the two main equations derived from fitting accurate measured data, namely, the "Del Grosso algorithm" or NRL II equation[3] and the UNESCO equation.[4] The former is based on absolute measurements with seawater limited to the salinity range 30%–41%. The latter, though determined over a broader range of salinities, is based on measurements by comparison with pure water but uses an incorrect equation for calculating the speed of sound in pure water under pressure. In all, the attempts to provide a universal accurate equation suffered from the lack of accurate measurements under pressure in both pure water and low-saline seawater. Recently however, new light was cast on the subject by an equation for sound speed in pure water under pressure based on new measurements up to 60 MPa, published by Belogol'skii *et al.*[5] These days, those wishing to calculate sound speed may attempt to implement the suitable equations in their own software, and there have also been attempts to provide online calculators.[6] However, the situation outlined above means that such attempts have to consider the two "complex" equations mentioned above plus an algorithm for calculating pressure from depth,[7,8] in addition to the two best available simple equations, those of Coppens[9] and Mackenzie.[10] Clearly, it would be more satisfactory to devise a single (and ideally simple) equation that gives essentially the same results as the best values obtained from data and equations in their relevant combinations of temperature T, salinity S, and pressure P calculated from depth (plus latitude Φ if necessary). At the same time, considering the fact that a new equation for pure water under pressure had been published, it is clearly of interest to see if it could not be used to replace the incorrect part of the UNESCO equation. It was then a pleasant surprise to discover that such a substitution could make the UNESCO equation be in much better agreement with Del Grosso's data for pressures up to some 30 MPa. Therefore, the use of this merged equation could provide reliable values for the miss-

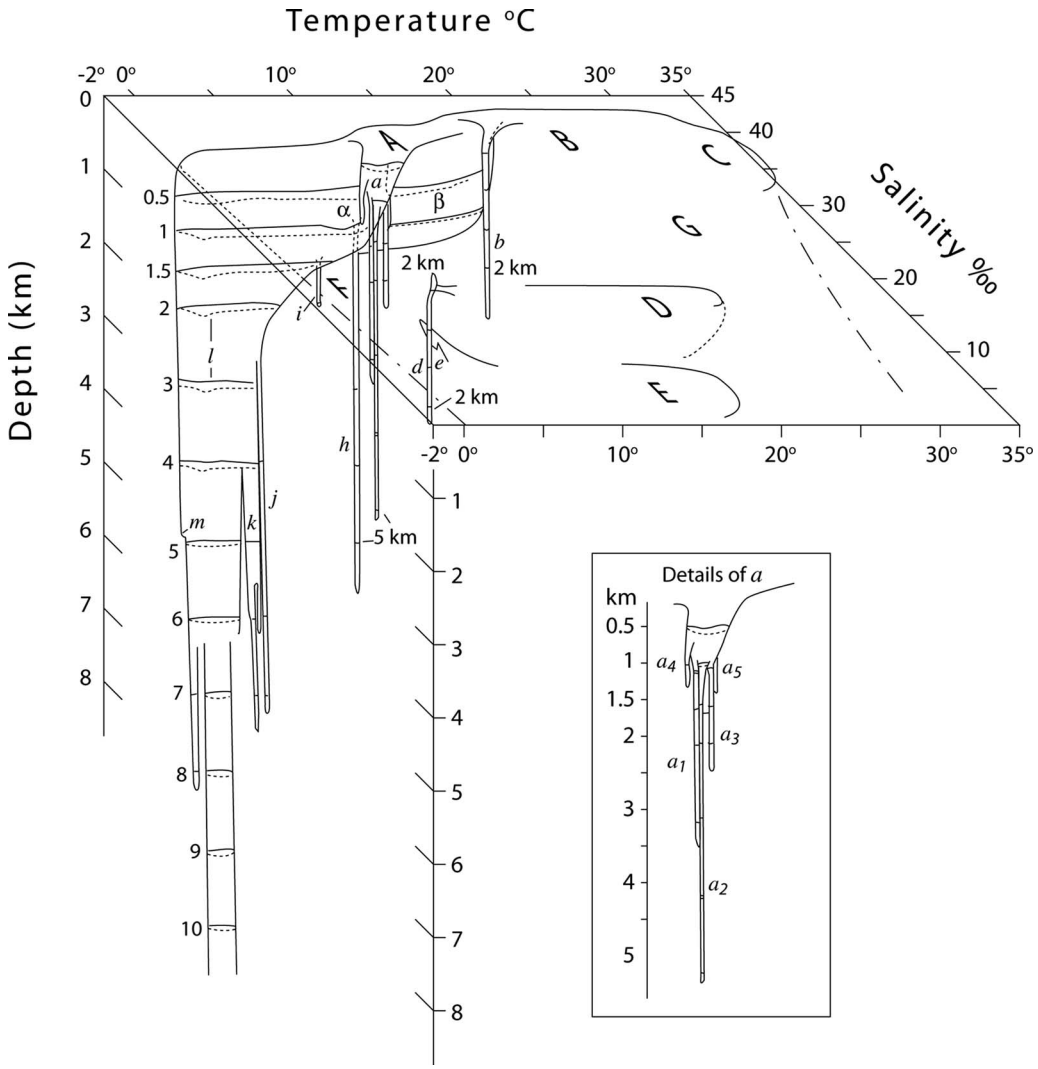


FIG. 1. Three dimensional representation of the temperature-salinity-depth volume of subspace enclosed by real Neptunian water.[11] The key is given Table I.

ing low-saline water under pressure found in the Baltic and the Black Sea, hence providing the possibility of developing a new equation that would be valid in all existing seawater conditions, from nearly pure water to salinities up to 42%. In view of the wide use of depth as a parameter in echo sounding, telemetry, acoustic tomography, ray tracing, and sonar applications, it was decided to use depth and latitude instead of pressure. An equation has thus been developed, which provides an accuracy always better than ± 0.2 m/s with respect to the most appropriate complex equation derived from fitting measured data (either NRL II or the merged equation, according to the required salinity range). The purpose of this paper is to present this equation and compare its results with those from the two reference equations, which are then discussed with respect to the accurate data. A comparison is also made with the equations of Coppens and Mackenzie. To make clear the motivation for the new equation, some discussion is first presented regarding the existing equations.

TABLE I. The key for Figure 1 showing the seas and oceans covered.

KEY	Surface water	Deep water	Water flowing into adjacent oceans
Mediterranean	A	*a*	α
Red Sea	B	*b*	β
Persian Gulf	C		
Black Sea	D	*d*	
Arctic coastal waters	F		
Tropical coast waters	G		
Sulu Sea		*h*	
Halmahera Basin		*i*	
"American Mediterranean"		*j*	
East Indian Basins		*k*	
Sea of Japan		*l*	
Arctic Ocean		*m*	

II. THE BASIC DATA AND THE EXISTING EQUATIONS

A. General

A valuable equation for sound speed in seawater should primarily fit the best available data for combinations of temperature, salinity, and pressure that exist in the oceans and seas. The subject of these existing *TSP* combinations in the oceans has already been examined by Leroy[11] who proposed

the terminology "Neptunian" waters to eliminate inland seas of extremely high salinities. These waters must, however, also eliminate the "hot spots" that exist in a few places at the bottom of various seas (Mediterranean, Red Sea, and Gulf of Mexico), where abnormal temperatures and salinities exist (>200%). In any case, the speed of sound has never been measured under such conditions, and if its knowledge was required, a separate formula should be developed from the results of specific measurements under these conditions. The reduced *TSP* domain of seawater of interest should be born in mind and is presented in Fig. 1 (see Table I), reproduced from Ref. 11.

In practice, the existing systematic, accurate, and reliable measurements of sound speed in seawater reduce to two sets only: those of Del Grosso and Mader (performed with Weissler or Mader),[12,13] leading to the NRL II equation, and those of Millero and Kubinski[14] and Chen and Millero,[15] leading to the UNESCO equation. A third series of earlier systematic measurements, leading to two now abandoned equations, had been performed by Wilson.[16,17] Unfortunately, it was found that some sets of his measurements were made at salinities or temperatures erroneously determined, and that his velocimeter was not reliable enough under high pressures. The remaining two main sets of data are quite different both in measuring methods and in contents and will be now examined, with comments on the resulting equations.

B. Del Grosso and the NRL II equation

All the measurements of Del Grosso were absolute measurements of sound speed performed with utmost care using extremely accurate laboratory sound velocimeters especially designed and built by himself and his team. In particular, the design of his pressure velocimeter allowed for the suppression of all problems due to the unavoidable deformations of the equipment under high pressure. For a greater accuracy, all Del Grosso's measurements at any nominal *TSP* combination were repeated (between four and seven times in almost all cases), temperature inside the acoustic cell being each time measured within ±0.001 °C. The explored *TSP* combinations almost completely covered the regular Neptunian domain down to depths of 9000 m, but limited to $30<S<41$ at $Z<2000$, and to $33<S<38$ and $0<T<15$ at $Z>2000$ (salinity quoted in percent, depth in meters, and temperature in °C). The conditions encountered in the Baltic, the Black Sea, and the Red Sea were therefore excluded. In all, these measurements covered 110 different *TSP* combinations, with 627 individual measurements (data points). Figure 2(a) gives an example of the distribution of these measurements.

Anderson then established the NRL II equation by taking the Del Grosso data points and adding seven additional values calculated at 0, 5, 10, 15, 20, 25, and 30 °C and $S=0\%$ from the "148 point" equation developed by Del Grosso and Mader for pure water at atmospheric pressure,[18] each with a weight of 5 (i.e., the equivalent of 35 data points). The 1968 international temperature scale was used, and the kg/cm^2 was the unit for pressure. This equation was subsequently readjusted by Wong and Zhu[19] for use with *T* values using the 1990 universal temperature scale and *P* values given in bars.

C. Chen and Millero and the UNESCO equation

In contrast to Del Grosso's measurements, those of Chen and Millero[15] (preceded by some from Millero and Kubinsky[14] at atmospheric pressure) were differential measurements performed with a commercial velocimeter and with hardly any repeated measurements. To calibrate the velocimeter, pure water was used at a range of pressures and temperatures, the latter being assumed to be the setting of the automatic controller (no independent measurements were made). Comparative measurements of velocity in seawater of various salinities were then made at similar temperature settings and pressures. The only repeated measurements were duplications and were only performed in four cases. In all, 340 different *TSP* combinations were investigated with a total of 375 data points. The *TSP* domain covered by these data is a "cubic" one, with temperatures between 0 to 40 °C, salinities between 5% and 40%, and pressures regularly spaced in 10 MPa intervals from 0 to 100 MPa, about half of these data corresponding to seawater conditions which do not exist on earth. Figure 2(b) gives an example of the distribution of these measurements for values around the same pressure as for Del Grosso and Mader, illustrating the great difference between the two approaches.

The set of data thus obtained was fitted by a polynomial equation giving the sound speed difference, ΔC, between saline and pure water as a function of *T*, *S*, and *P* as follows: $\Delta C=\{C(S,T,P)-C(0,T,P)\}$. To obtain from this an equation for sound speed in seawater, an equation giving the absolute values for pure water was required. There existed at the time only two such equations, one from Barlow and Yazgan[20] and one from Wilson.[21] (Del Grosso had not been given the time to make measurements in pure water under pressure.[22]) Barlow and Yazgan's equation was rejected because most temperatures considered were out of the domain of interest (a total of 8 values between 16 and 94 °C), and that of Wilson could not be used because evidence had proved that his data were unreliable. Chen and Millero chose to correct the data from Wilson, assuming that their departures from a smooth variation were due to errors in the temperature measurements or sound speed shifts specific to each temperature, but disregarding possible errors in pressure measurements or due to pressure effects.[23] This re-evaluation of Wilson's data led Chen and Millero to a new equation for sound speed in pure water and hence to an equation for seawater, later referred to as the UNESCO equation. This equation used the 1968 temperature scale and the bar as the unit for pressure. It was later recalculated by Wong and Zhu[19] for use with the 1990 temperature scale.

D. The discrepancies between the NRL II and the UNESCO equations

It is easy to observe that the NRL II and the UNESCO equations give different results for seawater under pressure. As the latter was intended to have a larger coverage includ-

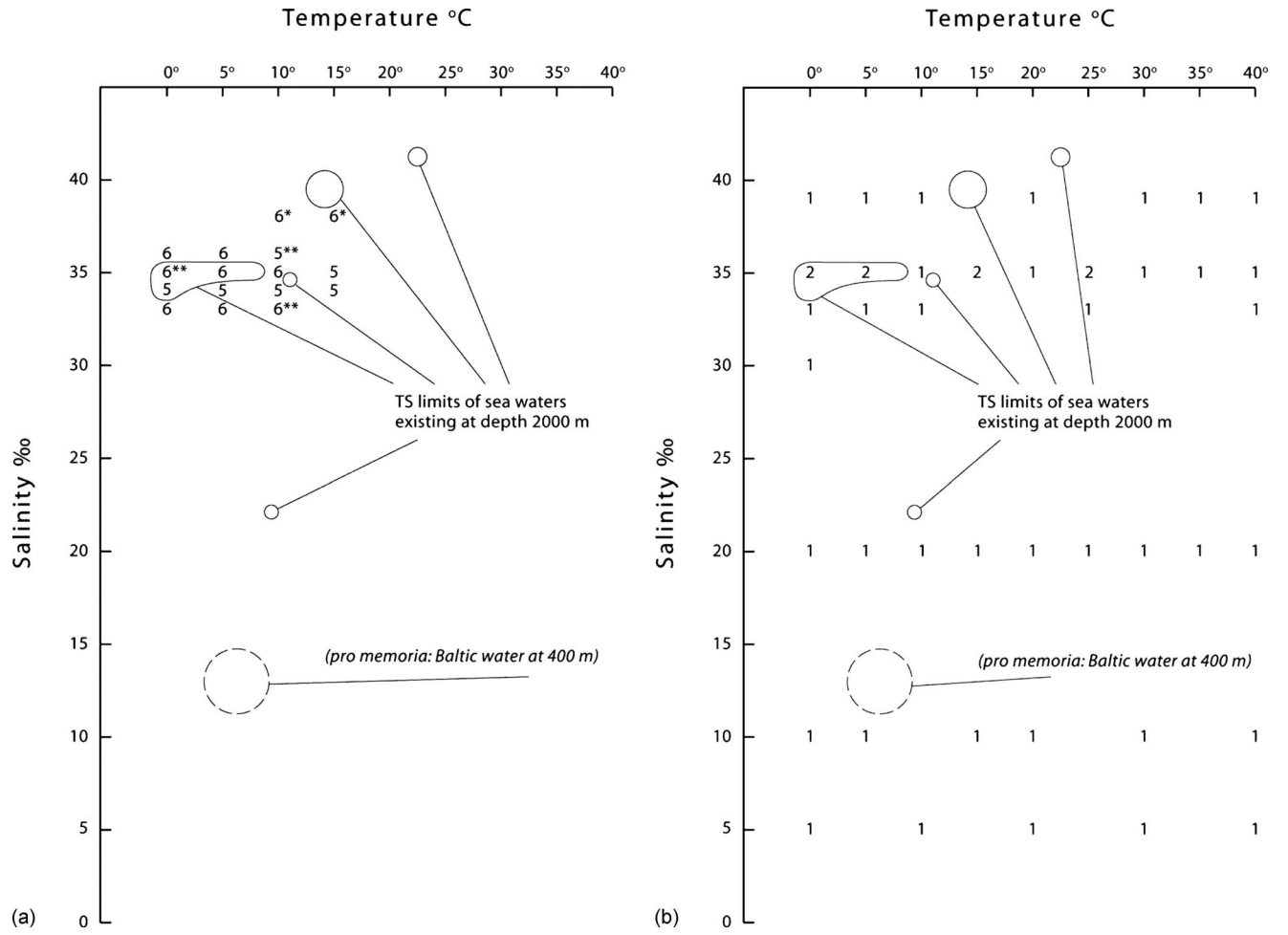


FIG. 2. (a) Distribution of the data points of Del Grosso and Mader as a function of T and S for comparison with existing TSP combinations. The values shown indicate the number of data points at $P=3000$ psi (gauge), while those labeled with * are for $P=2000$ psi (gauge) and those labeled with ** are for $P=4000$ psi (gauge). (b) Distribution of the data points of Chen and Millero at around 20 MPa as a function of T and S for comparison with existing TSP combinations. The values shown indicate the data points at $P=199.67$ bars.

ing low-saline seawater under pressure, it could be preferred by some users, at least in the corresponding seas. The fact that the UNESCO equation gave incorrect values under high pressure was proved by a number of acoustic experiments involving sound paths reaching great depths, and it was found that better agreements between observations and calculations were obtained using the NRL II equation.[24,25] This could obviously be due to the fact that the equation for pure water obtained from the re-evaluated data of Wilson gave incorrect values. This was first stressed by Dushaw *et al.*[1] and later agreed upon by Millero himself in a paper with Li.[22] They pointed out that making corrections independent of pressure was responsible for the errors and proposed a correction to apply to their equation in order to approach the values obtained with NRL II. This proposal, however, has not been adopted for correcting the UNESCO equation. In all, reliable data of sound speed in pure water under pressure were first needed before new comparisons with Del Grosso and with sea measurements involving great depths. The equation for pure water under pressure of Belogol'skii *et al.*[5] casts new light on this aspect of the problem.

E. The equation and data of Belogol'skii *et al.*

The equation of Belogol'skii *et al.* is based on differential measurements (with respect to atmospheric pressure conditions) performed with distilled water in the temperature range 0.4 to 40 °C and at pressures from 0 to 60 MPa by a quite different method than previous approaches. An acoustically transparent vessel filled with pure water was placed in a tank between a transmitter and a receiver and was pressurized at various values. Travel time differences with respect to atmospheric pressure conditions were measured and used to obtain differential values of sound speed, knowing the dimensions of the vessel, and its deformation with pressure calculated from the elastic properties of its material. Details regarding the overall accuracy of the method are, however, limited. A total of 231 data points are cited, resulting from measurements at 11 different temperatures and 21 pressures including some duplications. This set was used to obtain a matrix form polynomial development giving the differential values of sound ΔC as a function of temperature and pressure, to which the universally adopted equation of Del Grosso for sound speed in pure water at atmospheric pressure[18] was added to obtain the final equation. The equa-

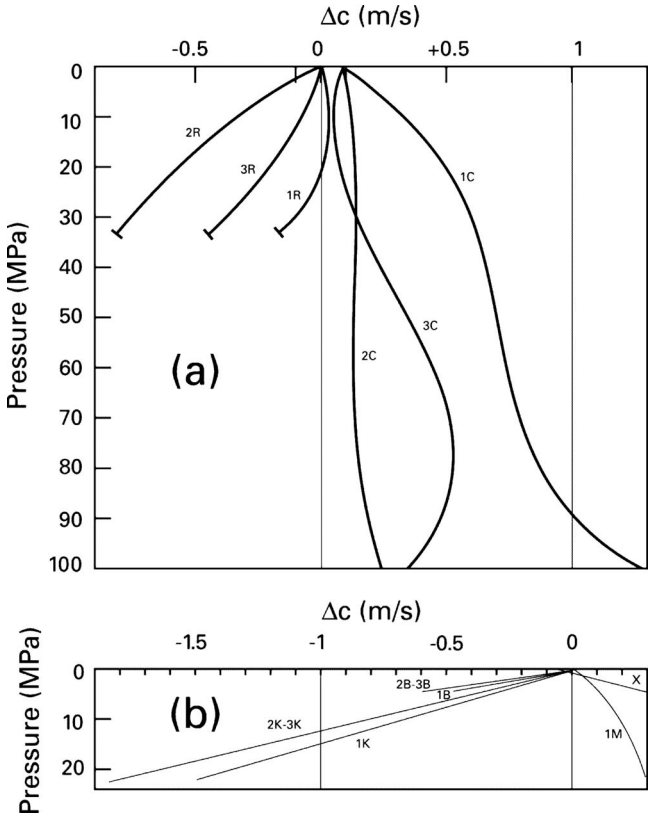


FIG. 3. Differences between sound speed values given by various equations with respect to the NRL II equation which is taken as the reference. The curves are numbered as follows: (1) UNESCO equation, (2) UNESCO equation as corrected by Millero and Li, and (3) UNESCO equation when replacing the pure water part of the equation by that of Belogol'skii *et al.* ("merged equation"). For the upper plot, (a) curves 1C, 2C, and 3C are the common oceans at depth (S=35%, T=2 °C); curves 1R, 2R, and 3R are for the Red Sea (S=40.5%, T=21.5 °C). The lower plot, (b), shows results for lower saline water, curves 1K, 2K, and 3K are for the Black Sea (S=22%, T=8.8 °C); curves 1B, 2B, and 3B are for the Baltic (S=13%, T=5 °C). In part (b) of the figure, two curves show the differences when taking the merged equation as the reference: curve 1M, UNESCO equation; curve X, NRLII.

tion for the differences ΔC is claimed to fit the original data with a standard deviation of 0.042 m/s and an arithmetic deviation smaller than 1 cm/s in absolute value.

F. A merged equation and its interest for low-saline water under pressure

Considering the accuracy claimed by the equation of Belogol'skii *et al.*, it is perfectly justifiable to try using it to replace the controversial part for pure water of the UNESCO equation, keeping the polynomial to be added for seawater unchanged. This gives rise to a merged equation for seawater, limited to pressures not exceeding 60 MPa. Using NRL II as the reference, Fig. 3 gives examples of the differences between the sound speeds obtained using three other equations: (i) the UNESCO equation, (ii) the UNESCO equation corrected according to Millero and Li, and (iii) the merged equation just described. One can observe that, in common cases, the departure of the UNESCO equation from NRL II grows markedly with increasing pressure, in particular, in the first 30 MPa. Such a behavior is not found with the merged equation which agrees rather well with NRL II until about 30 MPa. Overall, the results obtained with this merged equation are not entirely convincing, but provide an approximation to NRL II more closely than does the UNESCO equation. It can also be observed that the correction term of Millero and Li gives good results in common oceans but completely fails at high salinities under pressure (as in the case of the Red Sea). Therefore, for the work reported here, the results obtained with NRL II were chosen as the reference for *TSP* combinations encountered in the open oceans and seas with salinities higher than 30%.

The situation is very different with low-saline seawater under pressure. Del Grosso made no measurements at the so-called intermediate salinities, those between 0% and 30%. At atmospheric pressure, however, he had made measurements with distilled water leading to an equation that is universally accepted. As these values were included in the set of data fitted by NRL II, this equation gives good results at intermediate salinities at atmospheric pressure and is in good agreement with the UNESCO equation. However, as soon as these low-saline waters are placed under pressure, NRL II gives erroneous results because the equation is then employed in what corresponds to an extrapolation far outside the domain of the measurements. Figure 3(b) shows the departures between the equations at the low salinity *TSP* combinations found in the Black Sea and the Baltic. As pressure never exceeds 23 MPa in such waters, the use of the merged equation as the reference is then plainly justifiable. On Fig. 3(b), the curves marked "X" and "1M" are drawn using this reference instead of NRL II. The UNESCO equation is in good agreement, while Del Grosso's algorithm gives unacceptable results (outlined on the figure).

III. DEVELOPMENT OF THE NEW EQUATION

Following the observations presented above, it was decided to develop an equation that would give results in accordance with the Del Grosso algorithm in all oceans and seas, with the exception of the Black Sea and the Baltic where it should agree with the merged Belogol'skii, *et al.*-UNESCO equation. It was also decided to use the 1990 temperature scale, so that the equation would be consistent with Wong and Zhu's revisions. As many oceanographic data may employ the 1968 temperature scale, the user of the equation should check the scale and apply an elementary transformation if necessary. In the temperature range of existing seawater, this transformation is simply given by

$$T_{90} = 0.99975 \cdot T_{68}, \tag{1}$$

as it can be found from Fig. 1 and Table 6 of Ref. 26.

The use of results from the two reference equations instead of the original data for establishing the new equation could be considered an unusual approach, but in the end it is plainly justifiable.

This important point will be examined in Sec. IV.

The equation was developed with the aim of fitting all sound speed values that could be found in the oceans and seas below a salinity of 42%. For example, at salinities between 34% and 35% and at 5 °C, sound speed values must be accurate at depths from 0 to 2000 m only, but at all lati-

tudes lower than 60° (open oceans). Instead, in the same salinity range and at $T=-1$ °C, the accuracy must be ensured up to a depth $Z=4800$ m, at latitudes higher than 60° (Arctic Ocean), while at 10 °C it must be ensured up 5600 m of depth, but at 8° of latitude only (Sulu basin). The development of the equation followed the above approach, there being no need for it to give precise values of sound speed at TSZ and Φ combinations where seawater does not exist. It was also necessary that the maximum departures between the new equation and the reference ones should not exceed a given value, and that the agreement with the original data—when they were inside the existing seawater domain—should be satisfactory. The target for the maximum adopted acceptable departure from reference equations was then set at ± 0.3 m/s, with the hope of reaching a better value. It was also decided to use coefficients with reduced values of significant figures whenever possible. This point will be commented below.

The development of the equation was made using computer software through successive approaches, exploring each time the effect of modifying the coefficients in order to adjust them and simplify them if possible. At each given TSZ and Φ combination, the program calculated the pressure in MPa according to Leroy–Parthiot algorithm,[7,8] converted it to the required unit (kg/cm^2 or bars), and calculated the speed of sound given by the reference equation (NRL II or merged equation). The speed of sound given by the new equation was then calculated and compared with the reference, and plots of the differences were thus obtained and examined. The study was carried out systematically. For example, in the common oceans, the exploration was made by steps of 1000 m at each given $TS\Phi$ combination, steps of 0.5 or 1 °C at each selected $SZ\Phi$ combination, etc. In other cases (very deep trenches or closed basins), the specific values of T,S,Z and Φ were used and the proper correction term of the Leroy–Parthiot algorithm employed for calculating pressure.

It has been said that "the age of simple equations is over," a statement based on the fact that any personal computer can handle complex equations.[2] However, the analysis of the first results of the research, starting with common cases plus low-saline water, showed that coefficients with reduced numbers of significant figures could indeed be used with acceptable accuracy, and a decision was taken to proceed on that basis. This was even reinforced when it was found that an accuracy better than ± 0.2 m/s could be achieved. Some help in the development of the equation was provided by the examination of NRL II and Mackenzie's equation.[10] For example, there is no T^2P term in NRL II but a T^3P term is present. Similarly, there is no TZ^2 term in Mackenzie's equation but a TZ^3 term. However, the introduction of latitude is necessary, though it can be done without the use of trigonometric functions.

The new proposed equation is, with temperature T in °C (1990 universal temperature scale), salinity S in%, depth Z in meters, and latitude Φ in degrees, shown in

$$
\begin{aligned}
c = {} & 1402.5 + 5T - 5.44 \times 10^{-2}T^2 + 2.1 \times 10^{-4}T^3 \\
& + 1.33S - 1.23 \times 10^{-2}ST + 8.7 \times 10^{-5}ST^2 \\
& + 1.56 \times 10^{-2}Z + 2.55 \times 10^{-7}Z^2 - 7.3 \times 10^{-12}Z^3 \\
& + 1.2 \times 10^{-6}Z(\Phi - 45) - 9.5 \times 10^{-13}TZ^3 \\
& + 3 \times 10^{-7}T^2Z + 1.43 \times 10^{-5}SZ. \qquad (2)
\end{aligned}
$$

IV. DISCUSSION

A. The use of reference equations for the adjustment

The use of the values given by reference equations instead of the original measured data points was found preferable for various reasons. First, each data point corresponds to a unique measured pressure, and that pressure can be found at various depths according to location. For example, a pressure of 6000 psi (gauge) can be found at 4053 m (Arctic Ocean) and at 4078 m (equatorial water), a difference of 23 m leading to a speed variation of about 0.35 m/s at similar temperature and salinity values. For this reason, classical computer calculations of coefficients—for instance, by a minimization of the standard deviation of the model-data differences—are not well adapted, since the equation needs to use two variables, Z and Φ, while the data have only one parameter P. Another point in favor of reference equations for the adjustment of the new one is the fact that at high pressure, data points are slightly outside the range of existing temperatures. At depths greater than 5000 m, seawater only exists between 1.5 and 2.5 °C, with only two exceptions (American Mediterranean and Sulu Sea). At such pressures, the measurements were only made at two nominal temperatures, viz., 0 and 5 °C. There was no real need for the new equation to accurately fit the data at these temperatures while the aim was its accuracy between 1.5 and 2.5 °C. In fact, the NRL II equation can give good values in that range because it "inherits" the nonlinear variation of speed with temperature from data at lower pressures.

The above arguments are only valid provided the reference equations be in almost perfect agreement with the data they are based on. This is the case with NRL II which is claimed to give a standard deviation of 0.05 m/s with all Del Grosso's data points corresponding to "realistic combinations of the parameters" TSP for seawater and to temperature values from 0 to 30 °C in pure water at atmospheric pressure. This remarkably good fitting is, however, merely mentioned in the original paper.[3] It only appears in the abstract, and no more comment is given thereafter, nor are values of the mean and maximum deviations with data provided either. In the development of the new equation reported here, calculations have been made of these values for a total of 523 seawater data points of Del Grosso. These corresponded to all his measurements at atmospheric pressure and those under pressure at salinities of 34% and 35%. At the other salinities, some measurements were not considered if the combination of temperature and pressure was not realistic. The mean deviation of NRL II was then found to be +0.001 m/s only, with maximum departures of +0.16 and −0.2 m/s, hence the justification to use it at salinities higher than 33%.

TABLE II. Comparison between Eq. (2) (C_1) and NRL II as recalculated by Wong and Zhu (C_0) for common oceans as a function of depth at three latitudes.

Z (m)	P (MPa)	C_1 (m/s)	C_0 (m/s)	C_1-C_0 (m/s)
Common oceans at $\Phi=0°$, $S=35\%$, $T=2$ °C				
0	0.00	1457.985	1458.010	−0.025
1000	10.06	1474.279	1474.300	−0.021
2000	20.19	1491.027	1491.044	−0.017
3000	30.35	1508.175	1508.181	−0.006
4000	40.57	1525.667	1525.661	0.006
5000	50.83	1543.449	1543.434	0.015
6000	61.14	1561.464	1561.447	0.017
7000	71.49	1579.659	1579.646	0.013
8000	81.89	1597.976	1597.976	0.000
9000	92.33	1616.363	1616.381	−0.018
10000	102.81	1634.762	1634.802	−0.040
11000	113.34	1653.120	1653.181	−0.061
12000	123.90	1671.380	1671.456	−0.076
Common oceans at $\Phi=30°$, $S=35\%$, $T=2$ °C				
0	0.00	1457.985	1458.010	−0.025
1000	10.08	1474.315	1474.322	−0.007
2000	20.21	1491.099	1491.089	0.010
3000	30.40	1508.283	1508.250	0.033
4000	40.63	1525.811	1525.755	0.056
5000	50.90	1543.629	1543.553	0.076
6000	61.22	1561.680	1561.591	0.089
7000	71.59	1579.911	1579.815	0.096
8000	82.00	1598.265	1598.171	0.094
9000	92.45	1616.687	1616.601	0.086
10000	102.95	1635.122	1635.046	0.076
11000	113.49	1653.516	1653.448	0.068
Common oceans at $\Phi=60°$, $S=35\%$, $T=2$ °C				
0	0.00	1457.985	1458.010	−0.025
1000	10.10	1474.351	1474.365	−0.014
2000	20.27	1491.171	1491.179	−0.008
3000	30.48	1508.391	1508.387	0.004
4000	40.73	1525.955	1525.942	0.013
5000	51.04	1543.809	1543.790	0.019
6000	61.39	1561.896	1561.879	0.017
7000	71.78	1580.162	1580.154	0.008
8000	82.22	1598.553	1598.560	−0.007

As for the merged equation, there are no "direct" data points for this since the measurements with seawater were differential ones with respect to pure water, and those in pure water under pressure were not communicated. However, considering the accuracy claimed for the equation of Belogol'skii *et al.*, one can attempt to "reconstitute" data points from the Chen–Millero and Millero–Kubinski published differential values, adding to them the speed of sound calculated from the equation at each temperature and pressure. This work has been done for each realistic combination of temperature, salinity, and pressure. Unfortunately, due to the high number of combinations outside the existing domain and to the fact that only a single value was given for each combination, this corresponds to 18 data points only. Using these data, the merged equation gives a standard error of 0.15 m/s, a mean deviation of +0.023 m/s, and maximum departures of +0.22 and −0.32 m/s.

TABLE III. Comparison between Eq. (2) (C_1) and NRL II as recalculated by Wong and Zhu (C_0) for common oceans at a latitude of 30° as a function of temperature and salinity for two depths.

T (°C)	C_1 (m/s)	C_0 (m/s)	C_1-C_0 (m/s)
Common oceans, $\Phi=30°$, $P=80$ MPa ($Z=7808.13$ m), $S=34.7\%$			
1.0	1590.335	1590.272	0.063
1.5	1592.335	1592.281	0.054
2.0	1594.311	1594.272	0.039
2.5	1596.262	1596.246	0.016
3.0	1598.190	1598.202	−0.012
Common oceans, $\Phi=30°$, $P=80$ MPa ($Z=7808.13$ m), $T=2$ °C			
S (‰)	C_1 (m/s)	C_0 (m/s)	C_1-C_0 (m/s)
33.5	1592.610	1592.80	−0.193
34.0	1593.318	1593.41	−0.095
34.5	1594.027	1594.03	0.001
35.0	1594.736	1594.64	0.093
35.5	1595.445	1595.26	0.182
Common oceans, $\Phi=30°$, $P=5$ MPa ($Z=497.12$ m), $S=35\%$			
T (°C)	C_1 (m/s)	C_0 (m/s)	C_1-C_0 (m/s)
−2	1447.762	1447.727	0.035
0	1457.107	1457.097	0.010
2	1466.043	1466.055	−0.012
4	1474.579	1474.608	−0.029
6	1482.725	1482.767	−0.042
8	1490.492	1490.542	−0.049
10	1497.890	1497.941	−0.051
12	1504.929	1504.975	−0.047
14	1511.618	1511.654	−0.035
16	1517.969	1517.986	−0.017
18	1523.990	1523.981	0.009
20	1529.692	1529.649	0.043
Common oceans, $\Phi=30°$, $P=5$ MPa ($Z=497.12$ m), $T=8$ °C			
S (‰)	C_1 (m/s)	C_0 (m/s)	C_1-C_0 (m/s)
33	1488.004	1488.066	−0.062
34	1489.248	1489.302	−0.053
35	1490.492	1490.542	−0.049
36	1491.737	1491.786	−0.049
37	1492.981	1493.034	−0.053

B. Comparison of the equation with the reference ones and with others

The comparison of the equation with the reference ones has been done systematically in order to cover all existing seawater conditions, with the only exception being the anomalous sea bottom hot spots with salinities far exceeding 42‰. As said before, the Leroy–Parthiot algorithm has been used for transforming depth and latitude into pressure, each time with the appropriate term accounting for the gravitational anomaly. Tables II–IV give examples of such calculations for deep oceans and seas, and Table V for calculations with Baltic and Black Sea waters. The maximum departure with the reference equations was verified to stay within ±0.2 m/s. This accuracy of the equation is fully illustrated in Fig. 4.

With respect to the existing equations that use depth instead of pressure, the authors believe that the new equation is of much broader use and is more accurate. Simple equations such as those of Medwin[27] or Coppens[9] are either not

TABLE IV. Comparison between Eq. (2) (C_1) and NRL II as recalculated by Wong and Zhu (C_0) as a function of depth for the Eastern Mediterranean, Sulu Sea, Red Sea, and Arctic basin.

Z (m)	P (MPa)	C_1 (m/s)	C_0 (m/s)	C_1-C_0 (m/s)
Eastern Mediterranean at $\Phi=35°$, $S=38.7\%$, $T=13.6$ °C				
0	0.00	1506.586	1506.546	0.040
500	5.05	1514.746	1514.844	−0.098
1000	10.10	1523.018	1523.189	−0.171
1500	15.17	1531.387	1531.578	−0.191
2000	20.25	1539.838	1540.010	−0.172
2500	25.35	1548.356	1548.483	−0.127
3000	30.45	1556.926	1556.996	−0.070
3500	35.56	1565.532	1565.548	−0.016
4000	40.69	1574.160	1574.137	0.023
4500	45.82	1582.794	1582.761	0.033
5000	50.97	1591.418	1591.420	−0.002
5500	56.13	1600.019	1600.110	−0.091
Sulu Sea at $\Phi=9°$, $S=34.5\%$, $T=10.3$ °C				
0	0.00	1490.291	1490.256	0.035
500	5.02	1498.393	1498.441	−0.048
1000	10.05	1506.611	1506.702	−0.091
1500	15.10	1514.930	1515.024	−0.094
2000	20.16	1523.338	1523.405	−0.067
2500	25.24	1531.823	1531.841	−0.018
3000	30.32	1540.370	1540.330	0.040
3500	35.42	1548.969	1548.869	0.100
4000	40.52	1557.605	1557.455	0.150
4500	45.64	1566.266	1566.086	0.180
5000	50.77	1574.940	1574.757	0.183
5500	55.90	1583.613	1583.466	0.147
6000	61.05	1592.272	1592.211	0.061
Red Sea at $\Phi=20°$, $S=40.5\%$, $T=21.5$ °C				
0	0.00	1531.724	1531.641	0.083
500	5.04	1539.928	1540.011	−0.083
1000	10.09	1548.239	1548.402	−0.163
1500	15.15	1556.636	1556.817	−0.181
2000	20.22	1565.098	1565.257	−0.159
2500	25.31	1573.604	1573.723	−0.119
3000	30.40	1582.134	1582.219	−0.085
3500	35.51	1590.667	1590.745	−0.078
Arctic basin at $\Phi=80°$, $S=35\%$, $T=-1$ °C				
0	0.00	1444.429	1444.466	−0.037
1000	10.13	1460.820	1460.797	0.023
2000	20.31	1477.684	1477.663	0.021
3000	30.54	1494.981	1495.007	−0.026
4000	40.82	1512.674	1512.768	−0.094
5000	51.14	1530.724	1530.885	−0.161
5500	56.32	1539.872	1540.057	−0.185

accurate enough or of limited use (depths not exceeding 4000 m for the latter). Instead, a detailed comparison with the equation of Mackenzie[10] has been made, as this equation is more accurate and recognized as the best of those using depth. Compared with NRL II, the equation of Mackenzie is found to give values within ±0.3 m/s in the common oceans at depths up to 8000 m, but it departs thereafter, reaching as much as +1.4 m/s at 10000 m and +2.3 m/s at 11000 m. In the Artic ocean, the departure reaches −0.74 m/s at its maximum depth of 4800 m, and in the Red Sea sound speed differences are between −0.52 and +0.33 m/s. As for the low-saline seas that were not taken into account, the departures of Mackenzie's equation from the merged reference equation are as follows: (i) in Baltic water from zero to −0.61 m/s (ii) in the Black Sea from −0.35 to +0.185 m/s. All in all, the new equation proposed here stays within ±0.2 m/s of the reference equations, providing enhanced accuracy.

TABLE V. Comparison between Eq. (2) (C_1) and the merged equation (C_0) as a function of depth for the Black Sea and Baltic Sea (three temperatures).

Z (m)	P (MPa)	C_1 (m/s)	C_0 (m/s)	C_1-C_0 (m/s)
Black Sea at $\Phi=43°$, $S=22\%$, $T=8.8$ °C				
0	0.00	1469.457	1469.523	−0.066
500	4.99	1477.487	1477.570	−0.083
1000	9.99	1485.632	1485.709	−0.077
1500	15.01	1493.881	1493.936	−0.055
2000	20.03	1502.223	1502.247	−0.024
2500	25.07	1510.645	1510.639	0.006
Baltic Sea at $\Phi=58°$, $S=12\%$, $T=4$ °C				
0	0.00	1437.029	1437.144	−0.115
250	2.48	1440.993	1441.017	−0.024
500	4.96	1444.988	1444.930	0.085
Baltic Sea at $\Phi=58°$, $S=13\%$, $T=5$ °C				
0	0.00	1442.685	1442.808	−0.123
250	2.48	1446.653	1446.704	−0.051
500	4.96	1450.652	1450.636	0.016
Baltic Sea at $\Phi=58°$, $S=14\%$, $T=6$ °C				
0	0.00	1448.218	1448.349	−0.131
250	2.48	1452.190	1452.265	−0.075
500	4.96	1456.193	1456.214	−0.021

V. CONCLUSION

A simple equation [Eq. (2)] is proposed for the calculation of sound speed in seawater as a function of temperature, salinity, depth, and latitude. The equation uses the 1990 universal temperature scale. An elementary transformation [Eq. (1)] may be used if input data are based on the 1968 temperature scale. The accuracy of the equation with respect to the Del Grosso algorithm at salinities greater than 30% and to a merged equation based on the UNESCO and the equations of Belogol'skii *et al.* at lower salinities is better than ±0.2 m/s anywhere in the oceans and seas, and whatever the depth down to the greatest abysses. This is illustrated in Fig. 4 and in Tables II–V. The only exceptions to the validity of the equation in existing seawater are abnormal waters of salinity much higher than 42% (some inland "seas" and hot brine spots at the bottom of some seas).

The equation is of pure polynomial form, with 14 terms and coefficients of only one to three significant figures. Compared with a complete calculation using the Leroy–Parthiot algorithm to transform depth into pressure (9 terms, coefficients of 1–6 significant figures, 11 different corrective terms) and either the NRL II equation (19 terms, all coefficients with 12 significant figures) or the merged equation (41 terms with coefficients of 3–10 significant figures), this is a

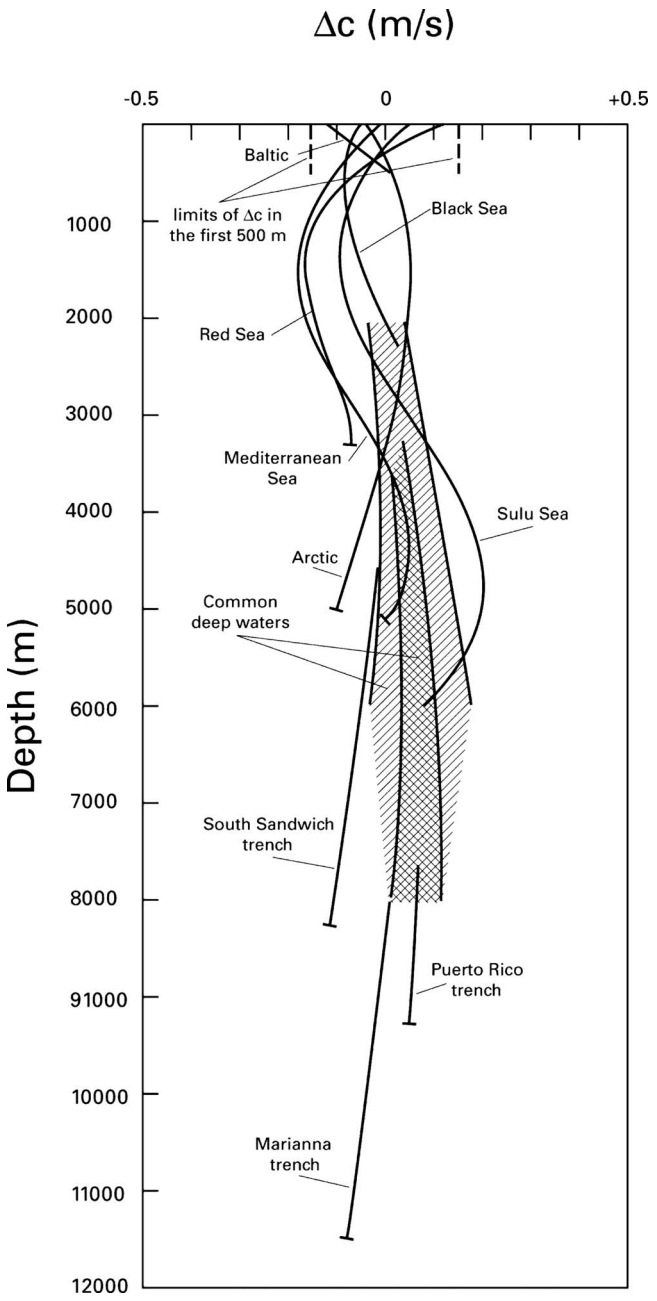


FIG. 4. Agreement of Eq. (2) with complex equations. Differences between the sound speed values (extreme cases) given by Eq. (2) and those obtained by transforming depth into pressure with the Leroy–Parthiot algorithm, and then using the NRLII equation (recalculated by Wong and Zhu). For the Black Sea and Baltic water (and those only), the reference was the merged equation.

substantial simplification. It is hoped that the equation will be welcomed by many of those who need to calculate sound speed in underwater acoustics applications.

ACKNOWLEDGMENTS

The authors would like to recognize the support of the National Measurement System Programmes Unit of the UK Department of Innovation Universities and Skills in this work.

[1]B. D. Dushaw, P. F. Forcester, B. D. Cornuelle, and B. M. Howe, "On equations for the speed of sound in seawater," J. Acoust. Soc. Am. **93**, 255–275 (1993).
[2]C. C. Leroy, "The speed of sound in pure and Neptunian water," *Handbook of Elastic Properties of Solids, Liquids, and Gases* (Academic, New York, 2001), Vol. IV, Chap. 2, pp. 23–81.
[3]V. A. Del Grosso, "New equation for the speed of sound in natural waters (with comparisons to other equations)," J. Acoust. Soc. Am. **56**, 1084–1091 (1974).
[4]N. P. Fofonoff and R. C. Millard, Jr., "Algorithm for computation of fundamental properties of seawater," UNESCO Technical Papers in Marine Science No. 44 (1983).
[5]V. A. Belogol'skii, S. S. Sekoyan, L. M. Samorukova, S. R. Stefanov, and V. I. Levtsov, "Pressure dependence of the sound velocity in distilled water," Meas. Tech. **42**, 406–413 (1999).
[6]Speed of Sound in Sea-Water, NPL, UK. Available at http://www.npl.co.uk/acoustics/techguides/soundseawater. Last viewed 6/5/2008.
[7]C. C. Leroy and F. Parthiot, "Depth-pressure relationships in the oceans and seas," J. Acoust. Soc. Am. **103**, 1346–1352 (1998).
[8]C. C. Leroy, "Depth-pressure relationships in the oceans and seas," J. Acoust. Soc. Am. **121**, 2447(E) (2007).
[9]A. B. Coppens, "Simple equations for the speed of sound in Neptunian waters," J. Acoust. Soc. Am. **69**, 862–863 (1981).
[10]K. V. Mackenzie, "Nine term equation for the sound speed in the oceans," J. Acoust. Soc. Am. **70**, 807–812 (1981).
[11]C. C. Leroy, "Development of simple equations for accurate and more realistic calculation of the speed of sound in seawater," J. Acoust. Soc. Am. **46**, 216–226 (1968).
[12]A. Weissler and V. A. Del Grosso, "Sound speed in pure water and seawater," J. Acoust. Soc. Am. **47**, 947–949 (1970).
[13]V. A. Del Grosso and C. W. Mader, "Speed of sound in sea-water samples," J. Acoust. Soc. Am. **52**, 961–974 (1972).
[14]F. J. Millero and T. Kubinski, "Speed of sound in seawater as a function of temperature and salinity at 1 atm," J. Acoust. Soc. Am. **57**, 312–319 (1975).
[15]C.-T. Chen and F. J. Millero, "Sound speed in seawater at high pressures," J. Acoust. Soc. Am. **62**, 1129–1135 (1977).
[16]W. D. Wilson, "Ultrasonic measurement of the velocity of sound in distilled and sea water," Naval Ordnance Report 6746, US Naval Ordnance Laboratory, White Oak, Maryland, 1960.
[17]W. D. Wilson, "Equations for the computation of the speed of sound in sea water," Naval Ordnance Report 6906, US Naval Ordnance Laboratory, White Oak, Maryland, 1962.
[18]V. A. Del Grosso and C. W. Mader, "Speed of sound in pure water," J. Acoust. Soc. Am. **52**, 1442–1446 (1972).
[19]G. S. K. Wong and S. Zhu, "Speed of sound in seawater as a function of salinity, temperature and pressure," J. Acoust. Soc. Am. **97**, 2235–2237 (1995).
[20]A. J. Barlow and E. Yazgan, "Pressure dependence of the velocity of sound in water as a function of temperature," J. Appl. Phys. **18**, 645–651 (1967).
[21]W. D. Wilson, "Speed of sound in distilled water as a function of temperature and pressure," J. Acoust. Soc. Am. **31**, 1067–1072 (1959).
[22]F. J. Millero and X. Li, "Comments on 'On equations for the speed of sound in seawater' [J. Acoust. Soc. Am. 93, 255–275 (1993)]," J. Acoust. Soc. Am. **95**, 2757–2759 (1994).
[23]C.-T. Chen and F. J. Millero, "Re-evaluation of Wilson's sound-speed measurements for pure water," J. Acoust. Soc. Am. **60**, 1270–1273 (1976).
[24]J. L. Spiesberger, "Is Del Grosso's sound-speed algorithm correct?," J. Acoust. Soc. Am. **93**, 2235–2237 (1993).
[25]C. S. Meinen and D. R. Watts, "Further evidence that the sound-speed algorithm of Del Grosso is more accurate than that of Chen and Millero," J. Acoust. Soc. Am. **102**, 2058–2062 (1997).
[26]H. Preston-Thomas, "The International Temperature Scale of 1990 (ITS-90)," Metrologia **77**, 3–10 (1990).
[27]H. Medwin, "Sound speed in water: A simple equation for realistic parameters," J. Acoust. Soc. Am. **58**, 1318–1319 (1975).

Observations of clustering inside oceanic bubble clouds and the effect on short-range acoustic propagation

Thomas C. Weber
Center for Coastal and Ocean Mapping, University of New Hampshire, Durham, New Hampshire 03824

(Received 11 January 2008; revised 20 June 2008; accepted 27 August 2008)

It has recently been shown [Weber, T. C. *et al.* (2007). "Acoustic propagation through clustered bubble clouds," IEEE J. Ocean. Eng. **32**, 513–523] that gas bubble clustering plays a role in determining the acoustic field characteristics of bubbly fluids. In particular, it has been shown that clustering changes the bubble-induced attenuation as well as the ping-to-ping variability in the acoustic field. The degree to which bubble clustering exists in nature, however, is unknown. This paper describes a method for quantifying bubble clustering using a high frequency (400 kHz) multibeam sonar, and reports on observations of near-surface bubble clustering during a storm (14.6 m/s wind speed) in the Gulf of Maine. The multibeam sonar data are analyzed to estimate the pair correlation function, a measure of bubble clustering. In order to account for clustering in the mean acoustic field, a modification to the effective medium wave number is made. With this modification, the multibeam sonar observations are used to predict the effect of clustering on the attenuation of the mean field for short-range propagation (1 m) at frequencies between 10 and 350 kHz. Results for this specific case show that clustering can cause the attenuation to change by 20%–80% over this frequency range. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2990707]

PACS number(s): 43.30.Ft [RCG] Pages: 2783–2792

I. INTRODUCTION

It has long been recognized that the propagation of acoustic waves in fluid media is changed in the presence of gas bubbles (Mallock, 1910; Wood, 1930), and that the performance of sonar systems can be adversely affected by increased sound attenuation and changes in sound speed (National Defense Research Committee, 1946). The theories that have been set forth and widely used to describe these changes treat the two-phase bubbly mixtures as a single "effective" medium with characteristics describing the average propagation (Mallock, 1910; Wood, 1930; Foldy, 1945; Van Wijngaarder, 1972; Commander and Prosperetti, 1989; Henyey, 1999). Regardless of the approach used when developing these theories (multiple scattering, the continuum approximation, and the diagram method), it is often assumed that the locations of the bubbles are statistically independent random variables.

Weber *et al.* (2007) explored the changes in acoustic field statistics in the presence of statistical dependence on the positions of the bubbles, a scenario which is described as bubble clustering after Shaw *et al.* (2002) and is analogous to the preferential concentrations of particles explored by Eaton and Fessler (1994). Utilizing the multiple scattering approach developed by Foldy (1945) and an artificial model for bubble clustering, it was found that for bubble clouds in which both single and double scattering were important, the presence of bubble clustering changes both the attenuation of the coherent pressure field and the variability of the pressure amplitude (note that the term double scattering is used here to refer to scattering chains that have one interaction with two different bubbles). Examining a simulated population of 100 μm bubbles at their resonance frequency, Weber *et al.* (2007) found that clustering caused a decrease in the attenuation and an increase in the variability from that which would be expected from a nonclustered bubble cloud.

In general, the bubble-induced attenuation for scattering chains containing up to two bubbles can be calculated by examining the average pressure field observed at some field point $\vec{r}$ due to some distant source of sound as follows:

$$\langle p(\vec{r})\rangle = p_0(\vec{r}) + \int_V \int s(a_i)p_0(\vec{r}_i)G(\vec{r},\vec{r}_i)n(\vec{r}_i,a_i)da_i d\vec{r}_i + \int_V \int_V \int \int s(a_{i1})s(a_{i2})p_0(\vec{r}_{i1})G(\vec{r}_{i1},\vec{r}_{i2})G(\vec{r},\vec{r}_{i2}) \times g(|\vec{r}_{i1}-\vec{r}_{i2}|)n(\vec{r}_{i1},a_{i1})n(\vec{r}_{i2},a_{i2})da_{i1}da_{i2}d\vec{r}_i d\vec{r}_i, \quad (1)$$

which is the sum of the bubble-free pressure p_0, a single scattering term (the double integral), and a double scattering term (the quadruple integral). Each bubble is assumed to have a complex scattering coefficient s, which is a function of the radius a. The propagation along each scattering path is described by the free-space Green's function G. The positions and sizes of the bubbles within the volume V are treated as random variables; the bubble size distribution $n(\vec{r},a)$ is proportional to the marginal probability density function describing the position and size of each bubble. In the event of bubble clustering, the positions of two bubbles will not be statistically independent. That is, the term clustering is used here to imply that, given the location of a single bubble, the probability of finding a second bubble nearby is higher than the probability of finding it far away. In Eq. (1), clustering is represented in the double scattering term by the use of the pair correlation function g, which is parametrized on the dis-

tance between the two bubbles in the double scattering chain. As described by Weber *et al.* (2007), this pair correlation term is used to relate the joint probability density function to the marginal probability density functions for each bubble as follows:

$$n(\vec{r}_{i1},\vec{r}_{i2}) = n(\vec{r}_{i1})n(\vec{r}_{i2})g(|\vec{r}_{i1}-\vec{r}_{i2}|). \quad (2)$$

If there is no clustering, then the locations of the bubbles are statistically independent and the pair correlation function g is equal to 1. When the bubbles are clustered, the pair correlation function will be greater than 1 at short separation distances, but should still be convergent to 1 at large distances where the clustering becomes uncorrelated. It is important to note that for higher orders of scattering, higher order correlation functions are needed: Eqs. (1) and (2) describe scattering chains including up to two bubbles, and hence, the pair correlation function is required; triple scattering would require the three point correlation function, and so on.

A natural extension of the work described by Weber *et al.* (2007), who utilized an artificial model of bubble clustering, is to quantify the degree of bubble clustering that can be found in either natural (e.g., under oceanic breaking waves) or man-made (e.g., in ship wakes) environments. The source of clustering may be irregularities in the process creating the bubbles (collapsing air cavities for breaking waves or cavitation sites/rates for ship propellers), or turbulence if there is a gradient in the mean spatial distribution of bubbles (as is the case in the ocean, with higher quantities of bubbles at the surface than at depth). Quantifying clustering in bubble clouds is best done by direct measurements of correlation functions, such as the pair correlation function g, which are nominally parametrized as a function of the separation distance between the bubbles. It seems reasonable to expect, however, that these correlation functions will change as the dynamics causing the bubble clustering change. For example, in bubble clouds under breaking waves, the correlation functions may change as the wind speed (and therefore wave height) changes, there may be a depth dependency that is related to the dissipation of wave energy, and so on.

This paper describes a method for extracting the pair correlation function from high frequency (400 kHz) multibeam sonar measurements (Sec. II), and presents observations of bubble clustering under oceanic breaking waves in a single wind-wave condition (Sec. III). These observations are taken from a measurement of opportunity when the ship on which the multibeam sonar was mounted was returning to port because of bad weather, traveling through open ocean breaking waves, and the bubble clouds beneath them during a storm with a sustained 14.6 m/s wind speed. A modification to the effective medium wave number that accounts for clustering is presented, and a prediction of the attenuation of the mean acoustic field is made for short-range propagation (1 m) at frequencies between 10 and 350 kHz (Sec. IV).

II. ESTIMATING THE PAIR CORRELATION FUNCTION

Historically, multibeam sonars have primarily been used for seafloor mapping (Medwin and Clay, 1988), but more recently have been used for mapping quantities of interest in the water column, including bubbles (Weber *et al.*, 2005), pelagic fishes (Mayer *et al.*, 2002; Gerlotto and Paramo, 2003), marine mammals (Benoit-Bird and Au, 2003), particulates (Jones, 2003), and turbulent microstructure (Trevorrow, 2005). Typically, these sonars are high frequency, narrowband active sonars mounted in a downward or sideward looking mode on a wide range of vessels (e.g., 10 m launches, icebreakers, and ocean-going research vessels). They most often utilize a Mills cross array configuration (Mills and Little, 1953) with orthogonal transmit and receive arrays that are used to achieve both high spatial resolution (as small as 0.5° beamwidth) and a large field of view, most typically resulting in a wide fan of beams that interrogate a narrow slice of the water column. Similar to conventional echo sounders, some multibeam sonars, such as the one used in this work, provide the capability to record backscatter from scatterers within the water column for each receive beam.

In some scenarios, the pair correlation function for bubbles, or other scatterers, can be estimated directly from the multibeam sonar backscatter measurements. Consider the backscattered pressures p_1 and p_2 observed in two different multibeam sonar resolution cells (or voxels) on a single ping (i.e., for a single configuration of the bubbles) as follows:

$$p_1(\vec{r}) = \sum_{i=N_1}^{N_2} s(a_i)p_0(\vec{r}_i)G(\vec{r},\vec{r}_i),$$

$$p_2(\vec{r}) = \sum_{i=N_3}^{N_4} s(a_i)p_0(\vec{r}_i)G(\vec{r},\vec{r}_i), \quad (3)$$

where the scattered pressure observed at the receiving array located at $\vec{r}$ is the sum of the scattered waves from each set of bubbles within the resolution cell (for N bubbles in the cloud, it is assumed that bubbles N_1–N_2 contribute to the backscattered pressure p_1, and N_3–N_4 contribute to p_2). The incident pressure p_0 at each bubble is that due to the acoustic field that would exist in the absence of any bubbles, $s(a_i)$ is the scattering coefficient for bubble i with radius a_i, and the scattered acoustic wave propagates from each bubble, located at $\vec{r}_i$ to the receiving array according to the free-space Green's function G. There are two important observations to make about Eq. (3): It is assumed that only single scattering is important, and the bubble populations in both resolution cells are assumed to be different. The correlation between the two observations can then be written as

$$\langle p_1(\vec{r})p_2(\vec{r})\rangle = \int_{V_i}\int_{V_j}\int\int s(a_i)s(a_j)p_0(\vec{r}_i)p_0(\vec{r}_j) \times G(\vec{r},\vec{r}_i)G(\vec{r},\vec{r}_j)n(a_i,a_j,\vec{r}_i,\vec{r}_j)da_ida_jd\vec{r}_id\vec{r}_j, \quad (4)$$

where $\langle\ \rangle$ indicates an expected moment. The integrals over volumes V_i and V_j correspond to the two resolution cells, and the bubble size distribution n is proportional to the joint probability density function ρ describing the position and size of each bubble as follows:

$$\rho(a_i,a_j,\vec{r}_i,\vec{r}_j) = (1/N)^2 n(a_i,a_j,\vec{r}_i,\vec{r}_j). \quad (5)$$

Assuming that the bubble sizes are independent of one another and not a function of position, and that the joint probability density function describing the locations of the bubbles at $\vec{r}_i$ and $\vec{r}_j$ is equal to the product of the pair correlation function g and the squared marginal probability density function (Weber *et al.*, 2007), then Eq. (4) can be written as

$$\langle p_1(\vec{r})p_2(\vec{r})\rangle = \int_{V_i}\int_{V_j}\int\int s(a_i)s(a_j)p_0(\vec{r}_i)p_0(\vec{r}_j)G(\vec{r},\vec{r}_i) \times G(\vec{r},\vec{r}_j)g(|\vec{r}_i-\vec{r}_j|)n(a_i,\vec{r}_i)n(a_j,\vec{r}_j)da_i da_j d\vec{r}_i d\vec{r}_j. \quad (6)$$

If it is further assumed that the pair correlation function is independent of the position over the volume integrals (which correspond to the resolution cells of the multibeam sonar), the correlation function can be written as

$$\langle p_1(\vec{r})p_2(\vec{r})\rangle = g(r)\int_{V_i}\int s(a_i)p_0(\vec{r}_i)G(\vec{r},\vec{r}_i)n(a_i,\vec{r}_i)da_i d\vec{r}_i \times \int_{V_j}\int s(a_j)p_0(\vec{r}_j)G(\vec{r},\vec{r}_j)n(a_j,\vec{r}_j)da_j d\vec{r}_j = g(r)\langle p_1(\vec{r})\rangle\langle p_2(\vec{r})\rangle. \quad (7)$$

Thus, the pair correlation function $g(r)$, where r is the distance separating the two resolution cells located at $\vec{r}_i$ and $\vec{r}_j$, is

$$g(r) = \frac{\langle p_1(\vec{r})p_2(\vec{r})\rangle}{\langle p_1(\vec{r})\rangle\langle p_2(\vec{r})\rangle}. \quad (8)$$

Here, p_1 and p_2 are proportional to the measurements directly recorded from the multibeam sonar. The constants of proportionality could be found by calibrating the sonar so that the digitized returns from the sonar can be converted exactly into backscattered pressures, but are actually not required here because they would simply cancel out in Eq. (8). Further, the constants of proportionality are not required to be the same for both p_1 and p_2 if, for example, the volume of the multibeam resolution cells were different. Thus, any calibration error that might propagate forward into the pair correlation estimate error is canceled out. It is also worth noting that completely uncorrelated backscattered pressures would have a pair correlation of 1 according to the definition given by Eq. (8).

III. OBSERVATIONS OF THE PAIR CORRELATION FUNCTION

A. Environmental conditions

The methodology for estimating the pair correlation function described in Sec. II has been applied to multibeam data collected in the Gulf of Maine approximately 150 km east of Cape Cod, MA on 5 October 2006 aboard the R/V Hugh R. Sharp. During this time, the R/V Sharp was transiting in a westerly direction at a speed over ground of approximately 4 m/s, en route to Woods Hole, MA due to the deterioration of weather conditions. Of particular interest are data collected between 11:30 and 13:30 UTC. The average wind speed observed on the ship's mast at a height of 17 m during this time period was 14.6 m/s with a standard deviation of 1.8 m/s, and the average wind direction was 14 °T with a standard deviation of 8° (Fig. 1).

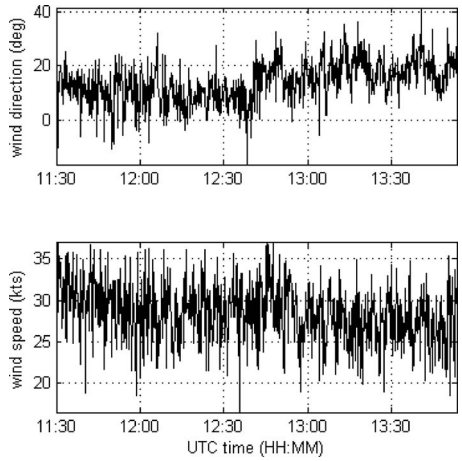


FIG. 1. The wind speed and direction during the multibeam sonar measurements on 5 October 2006.

Bubbles created under wind generated breaking waves are a ubiquitous feature of the ocean surface layer (Wu, 1981; Thorpe, 1982; Crawford and Farmer, 1987; Wu, 1998). The surface expression of a wave breaking event is a white cap, and it is estimated that for wind speeds of 10 m/s, approximately 1% of the surface will be covered by whitecaps (Monahan and Lu, 1990). Underneath these white caps are bubble clouds with void fractions of gas that have been observed to be as high as 10^{-3} (Terrill and Melville, 1997). Some attempts have been made to parametrize the bubble size distributions created in these circumstances as a function of wind speed and depth in the water. The parametrization that will be used here is from Novarini *et al.* (1998), which is based on the work done by both Monahan and Lu (1990) and Wu (1994).

The evolution of a bubble cloud can be separated into different stages: the first second or so after breaking where the void fraction of air can reach as high as 10^{-1}, the subsequent few seconds after breaking where the bubble plume is being advected downward by the jet associated with the breaking wave with void fractions as high as 10^{-3}–10^{-4}, the final stage where the breaking wave energy has largely decayed and the cloud of bubbles is more strongly affected by local currents and processes such as Langmuir circulation, where void fractions are $O(10^{-6})$, and the eventual decay into a background population of bubbles that has a very long lifetime (Monahan and Lu, 1990; Novarini *et al.*, 1998). The bubble clouds associated with a particular breaking wave have been defined as α-, β-, and γ-plumes, reflecting the changes in the cloud as it ages to 1 s, 3–4 s, and beyond (Monahan and Lu, 1990). Novarini *et al.* (1998) have parametrized the bubble size distribution and void fraction as a function of wind speed and depth, using a cubic wind speed dependence and an exponential decay in depth for the void

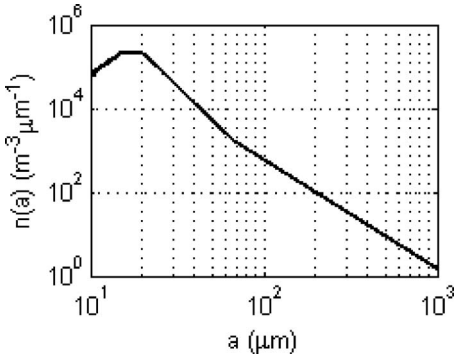


FIG. 2. The bubble size distribution for a β-plume at a wind speed of 14.6 m/s and a depth of 6 m using the parametrization given by Novarini *et al.* (1998).

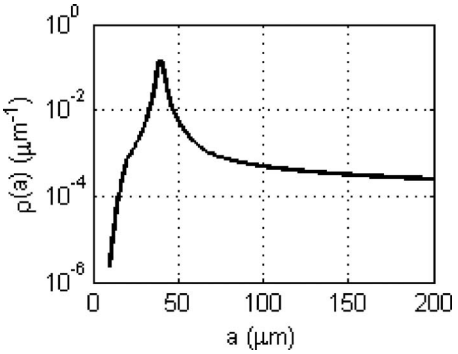


FIG. 4. The relative contributions of different bubble sizes at 400 kHz for bubble distributed as shown in Fig. 2.

fraction. Using this parametrization with a wind speed of 14.6 m/s and a depth of 6 m results in a void fraction of 4.5×10^{-6} and the bubble size distribution shown in Fig. 2.

B. Multibeam sonar measurements

Measurements of bubble clouds during the R/V Sharp's transit to Woods Hole, MA were made with a Reson 7125 SeaBat multibeam sonar. This sonar was mounted at a depth of approximately 4 m, looking downward, in one of the transducer bays in the retractable keel of the R/V Sharp. Gated cw pulses at a frequency of 400 kHz were transmitted from a linear transducer array oriented to generate a beam pattern that is approximately 1° wide at the half-power points (−3 dB) in the fore-aft direction and 155° wide in the athwart-ship direction. Backscattered acoustic signals were received on a second linear receiving array that is oriented orthogonally to the transmit array to form a Mills cross (Mills and Little, 1953) that can be steered to form 256 beams over a 128° sector of a plane. The receive beam width at broadside is nominally 0.5° wide in the athwart-ship direction, resulting in an effective beam pattern (transmit and receive arrays combined) that is $1°\times0.5°$. The native sampling scheme for the Reson 7125 produces pairs of 16 bit integers representing amplitude and phase for each sample in each beam. Each beam is sampled at a rate of approximately 34.5 ksamples/s. The transmitted pulse length for the data used here is 300 μs. Data were collected at a ping rate that varied between 5 and 10 pings/s.

An example of the acoustic backscatter data collected from the multibeam is shown in Fig. 3. These data correspond to a "snapshot" in time of the bubble clouds: The pulse travel times for the data shown in Fig. 3 range from 0 to 23 ms. The regions corresponding to higher backscatter levels are areas in which there are bubbles. The radius of a resonant bubble at 400 kHz at a depth of 6 m is approximately 10 μm using the simplified expression for resonance frequency given by Clay and Medwin (1977, p. 197), which is at the small size limit for the bubble size distribution shown in Fig. 2. At such high frequencies, however, larger off-resonance bubbles can make an appreciable contribution to the backscatter. Dahl (2001) suggested a useful metric for evaluating the relative contribution of a distribution of bubble sizes to backscatter at a single frequency by defining the following density function $\rho(a_s)$:

$$\rho(a_s)=\frac{\sigma_{\mathrm{bs}}(a_s)n(a_s)}{\int\sigma_{\mathrm{bs}}(a)n(a)da}, \tag{9}$$

where a_s is the radius of the bubble contributing to the backscatter and n is the bubble size distribution. The quantity $\sigma_{\mathrm{bs}}(a)$ is the backscattering cross section for a single bubble of radius a (Clay and Medwin, 1977) as follows:

$$\sigma_{\mathrm{bs}}(a)=\frac{a^2}{[(f_0/f)^2-1]^2+\delta^2}, \tag{10}$$

where f_0 is the resonance frequency of a bubble with radius a, f is the frequency for the incident acoustic wave, and δ is the total damping coefficient accounting for viscous losses, thermal losses, and reradiation. Equation (9) is shown in Fig. 4, and shows that the 400 kHz measurement will be dominated by bubbles with radii near 40 μm, four times larger than the resonant bubble size, with bubbles whose size is between 35 and 45 μm contributing to 68% of the overall scattering cross section.

It is important to note that one of the limitations of this experiment was that occasional dense bubble clouds would sweep directly underneath the multibeam transducers, rendering the measurements effectively useless. These instances are thought to represent very dense bubble clouds such as those that would be found directly under newly formed breaking waves (i.e., α-plumes) Such instances were manually removed from the data that follow. Although this makes the multibeam measurements less than optimal, creating a potential bias in any of the results due to human intervention in the data, enough measurements (more than 7000 pings)

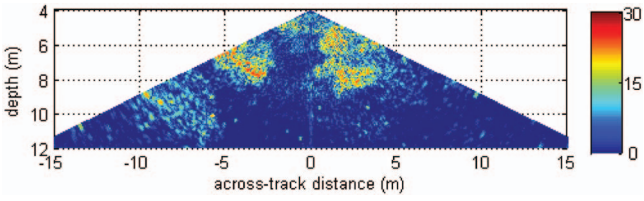


FIG. 3. (Color) An example of the multibeam data collected during the experiment. The data represent the combined 256 beams equally spaced over ±64°. The color scale represents relative intensity in decibels.

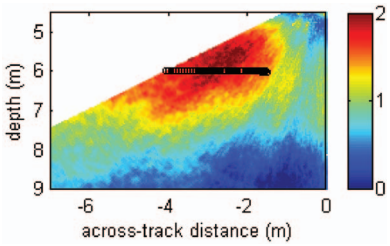


FIG. 5. (Color) The average backscattered intensity (proportional to the bubble number density) for the 365 pings (70 distinct bubble clouds) used in this work. The relative color scale is logarithmic. The black circles at a depth of 6 m show the individual sampling locations used to estimate the pair correlation function.

have been retained in order to make these data statistically significant of, at least, the less dense bubble clouds.

C. Pair correlation function estimate

One of the basic assumptions inherent in both Eq. (8) and in the treatment of propagation described by Weber *et al.* (2007) is that of statistical stationarity. Stationary data records can be difficult to achieve with transient forcing events such as breaking waves, where the bubble cloud appears at random locations within the multibeam field of view, and where the bubble number density is expected to change rapidly in time. For the data discussed here, this issue is exacerbated by the ship traveling through the water at 4 m/s: Bubble clouds may be encountered at any stage of their evolution. To resolve this problem, the approach taken here is to examine multiple bubble clouds, identify both a center of mass and a measure of the void fraction for each cloud, and to then conduct the analysis described in Sec. II for clouds that have a similar location and void fraction. Cloud centers are estimated by calculating both a weighted mean beam number $\bar{b}$ and range $\bar{r}$ as follows:

$$\bar{b} = \frac{\Sigma_{i=1}^{N} A_i^2 b_i}{\Sigma_{i=1}^{N} A_i^2}, \qquad \bar{r} = \frac{\Sigma_{i=1}^{M} A_i^2 r_i}{\Sigma_{i=1}^{M} A_i^2}. \tag{11}$$

The weights are the backscattered intensity, a quantity that is proportional to the scattering cross section and therefore to the bubble number density (Clay and Medwin, 1977, p. 203). A proxy for void fraction is estimated by summing the backscattered pressure magnitude squared over a specified region. In this work, the interest will be in bubble clouds whose center is located at an across-track range of -3.0 ± 1.0 m on the port side, and the region over which the void fraction is calculated extends for approximately 3 m in all directions away from the cloud center. With this focus in bubble cloud location, and including only those pings that have void fractions within $\pm 25\%$ of the average observed void fraction for clouds at this location, the resulting data set includes 365 pings over approximately 70 individual clouds. The average backscattered intensity for these pings is shown in Fig. 5.

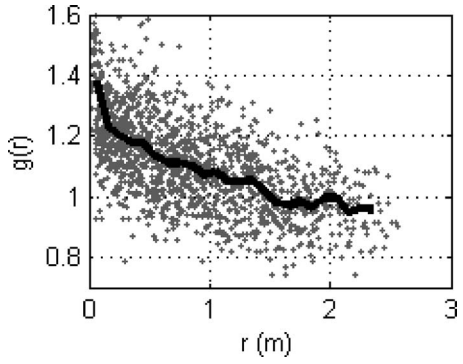


FIG. 6. Individual estimates of the pair correlation (gray dots) and the pair correlation function $g(r)$ plotted as a function of range r (black line).

The pair correlation function was calculated for these data using Eq. (8) at measurement locations whose depth was 6 m and across-track ranges were within 1.5 m of the cloud center (−3.0 m), as shown in Fig. 5. The ensemble averages in Eq. (8) correspond to an average over the subset of 365 pings. The measurement locations shown in Fig. 5 correspond to 55 different beams, resulting in 1485 individual estimates of the pair correlation at different lengths. These estimates are binned at 10 cm increments, according to the separation distance for each pair, and bin averages are computed in order to find a smoothed pair correlation function. Both the individual estimates and the smoothed pair correlation function are shown in Fig. 6.

The estimate for the pair correlation function shown in Fig. 6 shows an approximately linear decay with increasing range, reaching a maximum value of 1.25. The pair correlation function is greater than 1 at all ranges less than 1.5 m, meaning that given the location of a single bubble, it is more likely to find another bubble less than 1.5 m away than it is to find one that is greater than 1.5 m away. At ranges larger than 1.5 m, the pair correlation function is approximately equal to 1, meaning that there is no statistical dependency between the locations of bubbles separated by distances larger than 1.5 m. The pair correlation function shown in Fig. 6 appears to deviate from its linear behavior at the shortest ranges, less than ~5 cm. Pair correlation estimates at these short ranges are derived from adjacent beams, a situation that can introduce an anomalously high result due to overlapping beams. The beam pointing angles for the Reson 7125 are calculated such that their beam patterns will overlap at the half-power points (−3 dB). Thus, a significant portion of the sampling volume will be shared between adjacent beams, violating the assumption of unique bubble populations in the derivation of Eq. (8). The measurements corresponding to the shortest range (5 cm) are not used in the analysis that follows.

IV. THE EFFECT ON THE ACOUSTIC FIELD

Weber *et al.* (2007) numerically integrated Eq. (1) in order to calculate the effect of bubble clustering on the attenuation of the average field for a specific geometry and bubble cloud size, finding that clustering can cause the attenuation to diverge from the nonclustered case above some

bubble number density. Without a specific case study in mind (beyond the general case of bubble clouds under breaking waves), the approach here is to more generally analyze the effect of clustering on attenuation. This is done by first expanding Eq. (1) for the case of *nonclustered* bubbles to include the triple scattering term

$$\begin{aligned}\langle p(\vec{r})\rangle &= p_0(\vec{r}) + S\int_V p_0(\vec{r}_i)G(\vec{r},\vec{r}_i)n(\vec{r}_i)d\vec{r}_i + S^2\int_V\int_V p_0(\vec{r}_{i1})G(\vec{r}_{i1},\vec{r}_{i2})G(\vec{r},\vec{r}_{i2})n(\vec{r}_{i1})n(\vec{r}_{i2})d\vec{r}_{i1}d\vec{r}_{i2} \\ &+ S^3\int_V\int_V\int_V p_0(\vec{r}_{i1})G(\vec{r}_{i1},\vec{r}_{i2})G(\vec{r}_{i2},\vec{r}_{i3})G(\vec{r},\vec{r}_{i3})n(\vec{r}_{i1})n(\vec{r}_{i2})n(\vec{r}_{i3})d\vec{r}_{i1}d\vec{r}_{i2}d\vec{r}_{i3i} + \cdots \\ &= p_0(\vec{r}) + p_1(\vec{r}) + p_2(\vec{r}) + p_3(\vec{r}) + \cdots , \end{aligned} \tag{12}$$

where the ellipsis indicates that this iterated multiple scattering series carries on *ad infinitum*. The term S in Eq. (12) is a shorthand notation for the integral of the scattering coefficient s weighted by the bubble size distribution $n(a)$ over all bubble sizes a, which is assumed here to be separable from the integrals over position. A shorthand notation of each scattering term has also been introduced, where p_1 represents the single scattered pressure, p_2 represents the double scattered pressure, and p_3 represents the tripled scattered pressure.

Equation (12) can be reduced by applying the Helmholtz operator $(\nabla^2+k_0^2)$ to both sides of the equation in a manner similar to that done by Foldy (1945) as follows:

$$\begin{aligned}(\nabla^2 + k_0^2)\langle p(\vec{r})\rangle &= 0 - 4\pi S - 4\pi S^2\int_V p_0(\vec{r}_i)G(\vec{r},\vec{r}_i)n(\vec{r}_i)\mathrm{d}\vec{r}_i - 4\pi S^3\int_V\int_V p_0(\vec{r}_{i1})G(\vec{r}_{i1},\vec{r}_{i2})G(\vec{r},\vec{r}_{i2})n(\vec{r}_{i1})n(\vec{r}_{i2})d\vec{r}_{i1}d\vec{r}_{i2} - \cdots \\ &= -4\pi S[p_0(\vec{r}) + p_1(\vec{r}) + p_2(\vec{r}) + \cdots] = -4\pi S\langle p(\vec{r})\rangle . \end{aligned} \tag{13}$$

Equation (13) is identical in result, and, essentially, method of derivation, to the result given by Foldy (1945), and can be solved to find an effective medium wave number k as follows:

$$k^2 = k_0^2 + 4\pi S = k_0^2 + 4\pi\int s(a)n(a)da. \tag{14}$$

When clustering is present, Eq. (12) takes the form

$$\begin{aligned}\langle p(\vec{r})\rangle &= p_0(\vec{r}) + S\int_V p_0(\vec{r}_i)G(\vec{r},\vec{r}_i)n(\vec{r}_i)d\vec{r}_i + S^2\int_V\int_V p_0(\vec{r}_{i1})G(\vec{r}_{i1},\vec{r}_{i2})G(\vec{r},\vec{r}_{i2})g(|\vec{r}_{i1}-\vec{r}_{i2}|)n(\vec{r}_{i1})n(\vec{r}_{i2})d\vec{r}_{i1}d\vec{r}_{i2} \\ &+ S^3\int_V\int_V\int_V p_0(\vec{r}_{i1})G(\vec{r}_{i1},\vec{r}_{i2})G(\vec{r}_{i2},\vec{r}_{i3})G(\vec{r},\vec{r}_{i3})g(|\vec{r}_{i1}-\vec{r}_{i2}|,|\vec{r}_{i2}-\vec{r}_{i3}|)n(\vec{r}_{i1})n(\vec{r}_{i2})n(\vec{r}_{i3})d\vec{r}_{i1}d\vec{r}_{i2}d\vec{r}_{i3i} + \cdots . \end{aligned} \tag{15}$$

The fourth term in Eq. (15) contains the triple correlation function $g(|\vec{r}_{i1}-\vec{r}_{i2}|,|\vec{r}_{i2}-\vec{r}_{i3}|)$ that is a function of the distances separating both the first two and the last two bubbles in a scattering chain. In some circumstances, this triple correlation function should be equivalent to the pair correlation function: This would occur, for example, when the joint probability density function (PDF) $n(\vec{r}_{i1},\vec{r}_{i2},\vec{r}_{i3})$ describing the locations of bubbles $i1$, $i2$, and $i3$ is equal to the product of the joint pdf $n(\vec{r}_{i1},\vec{r}_{i2})$ and the marginal pdf $n(\vec{r}_{i3})$. This scenario could be expected to occur when $i2$ and $i3$ are far enough apart to be statistically independent (greater than 1.5 m for the bubble clouds examined in Sec. III). For bubble clouds in which the size of the bubble cloud, as well as the volume of integration for the integrals in Eq. (15), is large, the triple correlation function would be approximately equal to the double correlation function for the majority of the integration domain.

Equation (15) can also be reduced by applying the Helmholtz operator to each of the terms separately as follows:

$$(\nabla^2 + k_0^2)\langle p(\vec{r})\rangle = 0 - 4\pi S - 4\pi S^2 \int_V p_0(\vec{r}_{i1}) G(\vec{r}_{i1},\vec{r}) g(|\vec{r}_{i1} - \vec{r}|) n(\vec{r}_{i1}) d\vec{r}_{i1} - 4\pi S^3 \int_V \int_V p_0(\vec{r}_{i1}) G(\vec{r}_{i1},\vec{r}_{i2}) G(\vec{r}_{i2},\vec{r}) g(|\vec{r}_{i1} - \vec{r}_{i2}|, |\vec{r}_{i2} - \vec{r}|) n(\vec{r}_{i1}) n(\vec{r}_{i2}) d\vec{r}_{i1} d\vec{r}_{i2} - \cdots . \quad (16)$$

Unlike Eq. (13), the application of the Helmholtz operator does not result in the reduction by one of the scattering chain hierarchies. Instead, correction terms must be included as follows:

$$(\nabla^2 + k_0^2)\langle p(\vec{r})\rangle = -4\pi S[p_0(\vec{r}) + p_1(\vec{r}) + p_2(\vec{r}) + \cdots] - 4\pi S^2 \int_V p_0(\vec{r}_{i1}) G(\vec{r}_{i1},\vec{r}) [g(|\vec{r}_{i1} - \vec{r}|) - 1] n(\vec{r}_{i1}) d\vec{r}_{i1} - 4\pi S^3 \int_V \int_V p_0(\vec{r}_{i1}) G(\vec{r}_{i1},\vec{r}_{i2}) G(\vec{r}_{i2},\vec{r}) [g(|\vec{r}_{i1} - \vec{r}_{i2}|, |\vec{r}_{i2} - \vec{r}|) - g(|\vec{r}_{i1} - \vec{r}|)] n(\vec{r}_{i1}) n(\vec{r}_{i2}) d\vec{r}_{i1} d\vec{r}_{i2} - \cdots . \quad (17)$$

The third term in Eq. (17), which is based on the difference between the triple and pair correlation functions, should vanish when the volume of integration is much larger than the length over which bubble positions are found to be statistically dependent. This is essentially the same argument made when using the quasicrystalline approximation (QCA) to account for the hole correction in very densely packed scatterers (Lax, 1952). When this occurs, Eq. (17) can be approximated as

$$(\nabla^2 + k_0^2)\langle p(\vec{r})\rangle \cong -4\pi S[p_0(\vec{r}) + p_1(\vec{r}) + p_2(\vec{r}) + \cdots] - 4\pi S^2 \times \int_V p_0(\vec{r}_{i1}) G(\vec{r}_{i1},\vec{r}) [g(|\vec{r}_{i1} - \vec{r}|) - 1] \times n(\vec{r}_{i1}) d\vec{r}_{i1}, \quad (18)$$

leading to a new effective wave number that accounts for clustering as follows:

$$k^2 \cong k_0^2 + 4\pi S + 4\pi S^2 \langle p(\vec{r})\rangle^{-1} \int_V p_0(\vec{r}_{i1}) \times G(\vec{r}_{i1},\vec{r}) [g(|\vec{r}_{i1} - \vec{r}|) - 1] n(\vec{r}_{i1}) d\vec{r}_{i1}. \quad (19)$$

There are several features of Eq. (19) that are worth pointing out. First, as the clustering grows weaker and the pair correlation function becomes close to 1, this solution converges to the nonclustered solution given in Eq. (14). Second, the correction for clustering grows both as the volume of bubbles grows and as the distance of propagation grows. For example, consider a plane wave that would propagate in the bubble-free medium as $\exp(-jk_0x)$. The effective wave number described by Eq. (19) would then be written as

$$k^2 \cong k_0^2 + 4\pi S + 4\pi S^2 \times \frac{\int_V \exp(-jk_0 x_{i1}) G(\vec{r}_{i1},\vec{r}) [g(|\vec{r}_{i1} - \vec{r}|) - 1] n(\vec{r}_{i1}) d\vec{r}_{i1}}{\exp(-jkx)}. \quad (20)$$

Considering a widely separated source and receiver with a bubble cloud in between, the correction term in Eq. (20) will grow as either the size and or density of the bubble cloud grows, thereby decreasing the magnitude of the term in the denominator. This is in qualitative agreement with result found by Weber *et al.* (2007), who found that the relative importance of double scattering increases with both the number of bubbles and the bubble cloud size.

Evaluating Eq. (20) is situational dependent, not only on the bubble cloud parameters but also on the source/receiver geometry. However, a simple example is provided here in which Eq. (20) is numerically evaluated. The effective wave number k is solved for iteratively because of its presence in both the left-hand side and in the denominator of the third term on the right-hand side of Eq. (20). Starting with an initial guess for the effective wave number k_g, which is used in the denominator on the right-hand side, Eq. (20) is evaluated in order to find an estimate for k. This new estimate is then used as the new k_g until the magnitude of the terms $\exp(-jk_g x)$ and $\exp(-jkx)$ are within 1% of each other. In this manner, Eq. (20) has been evaluated using the pair correlation function shown in Fig. 6, the bubble size distribution shown in Fig. 2, and assuming that the hydrophone [i.e., the field point $\vec{r}$ in Eq. (20)] observes incident plane waves at a position that is located 1 m from the edge of a bubble cloud such that the term in the denominator of Eq. (20) includes the attenuation $\exp(-\mathrm{Im}\{k\}1)$. The bubble cloud is assumed to be large so that the approximation (analogous to the QCA) leading to Eq. (20) is valid. The results shown in Fig. 7 describe a difference in attenuation between the clustered and nonclustered bubble clouds that varies between 20% and 80% over a frequency range between 10 and 350 kHz. The attenuation for the clustered bubble cloud is lower than that predicted for the nonclustered bubble cloud at frequencies between 10 and 230 kHz, and higher at frequencies between 230 and 350 kHz.

The reason for the transition from a lower than expected attenuation to a higher than expected attenuation in Fig. 7 is based on the competition between the last two terms in Eq. (20). Considering that the last two terms in Eq. (20) are small in comparison to k_0^2, the attenuation can be written as

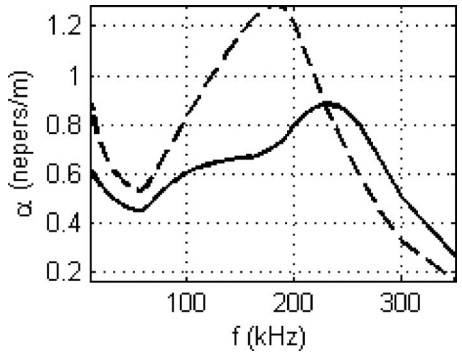


FIG. 7. The predicted attenuation for nonclustered [Eq. (14); dashed line] and clustered bubble clouds [Eq. (18), solid line].

$$\alpha \cong -\frac{2\pi}{k_0}\,\mathrm{Im}\left\{S+S^2 \times \frac{\int_V \exp(-jk_0x_{i1})G(\vec{r}_{i1},\vec{r})[g(|\vec{r}_{i1}-\vec{r}|)-1]n(\vec{r}_{i1})d\vec{r}_{i1}}{\exp(-jkx)}\right\}. \tag{21}$$

From Fig. 6, the pair correlation function can be approximated as $g=1.25-(0.25/1.25)r$, in which case the volume integral in Eq. (21) is proportional to $1/jk_0$ and the second term in the brackets of Eq. (21) is proportional to S^2/jk_0. The imaginary part of this latter quantity, $\mathrm{Im}\{S^2/jk_0\}$, has been plotted as a function of frequency in Fig. 8. Noting that $\mathrm{Im}\{S\}$ is negative at all frequencies (hence the positive attenuation in Fig. 7), clustering should cause the attenuation to decrease where $\mathrm{Im}\{S^2/jk_0\}$ is greater than zero and increase where $\mathrm{Im}\{S^2/jk_0\}$ is less than zero. A comparison of Figs. 7 and 8 shows that this is occurring here: The attenuation for the clustered bubble cloud becomes higher than that for the nonclustered bubble cloud at 230 kHz, where the quantity $\mathrm{Im}\{S^2/jk_0\}$ changes sign.

V. DISCUSSION

When using stochastic approaches for examining the effect of bubbles on the propagation of acoustic waves, it is often assumed that the positions of the bubbles are statistically independent random variables. This assumption has been tested here using multibeam sonar backscatter from bubbles under oceanic breaking waves in an open ocean condition at a wind speed of 14.6 m/s, and found to be not generally valid. The multibeam data indicate that the bubble positions were positively correlated (clustered) at ranges less than 1.5 m, and independent of one another at greater ranges. By itself, this finding does not mean that commonly used effective medium theories are in error. For scenarios where only single scattering is non-negligible, the important scattering paths contain an interaction with only one bubble and so the *average* of the scattered waves associated with these paths depends only on the marginal probability density function describing the location of individual bubbles. This might happen at lower wind speeds where the bubble number density is less. The correlations between the fluctuating positions of bubbles become important for the mean acoustic field when double scattering terms (or higher) become important.

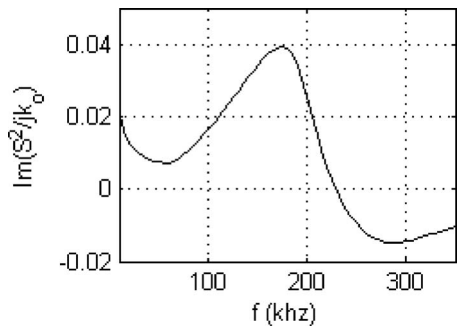


FIG. 8. A term proportional to the clustering induced modification of the attenuation plotted as a function of frequency.

A correction term for the average acoustic field that accounts for bubble clustering has been suggested in this paper. This correction term vanishes when the bubble positions are statistically independent (i.e., when the pair correlation function g is equal to 1), and, because it is based on a double scattering term in an iterated multiple scattering series, will become negligible when double scattering is negligible. The correction term used in this work is an approximation similar to the QCA described by Lax (1952). This approximation suggests that the clustering correction for double scattering dominates corrections for higher orders of scattering, and is expected to hold true when the positions of the bubbles are statistically dependent only on ranges much shorter than the size of the bubble cloud contributing to the mean acoustic field.

Using the pair correlation function estimated from the multibeam data, and a predicted bubble size distribution for oceanic breaking waves, it was found that the application of the correction term caused a substantial deviation in the mean acoustic field from that which would be predicted in the absence of clustering. Whether or not the clustered bubble cloud attenuation is higher or lower than the nonclustered bubble cloud attenuation will be a function of the frequency dependence of the bubble scattering coefficient integrated over the bubble size distribution. This means that the correction term will be a function of both the statistical dependency between the positions of the bubbles and the size distribution of the bubbles. It is important to note that the data described here, and their relationship to the importance of bubble clustering, can be considered only an example of the type of fluctuating spatial structures that exist within bubble clouds in the ambient ocean. It is quite likely that scenarios exist where the bubble clustering (and associated correlation functions) are either stronger or weaker, and presumably there is a relationship between the observed correlations and the bubble cloud forcing functions (e.g., different wind speeds may mean different correlation functions).

The deviations in the mean acoustic field were found to exist over a wide frequency range, from 10 to 350 kHz. However, the implicit assumptions in this analysis were that the clustering observed with the 400 kHz multibeam was the same across all bubble sizes, and that the overall bubble size distribution was adequately described by a model of this distribution for a β-plume. In the absence of a multifrequency multibeam sonar and direct measurements of the bubble size

distribution, these assumptions make the problem tractable. There is a possibility that these assumptions may not have been appropriate. For example, the assumption that all bubbles respond to the fluid flow in the same way and exhibit identical clustering may not hold true for all bubble sizes. A separate body of research describing the clustering of particles in turbulent flow fields (e.g., Eaton and Fessler, 1994) suggests that not all bubbles will respond to the flow fields generating the clustering equally. This research indicates that the degree to which bubble size sorting occurs in the ocean should depend on the Stokes number (defined as the ratio of the bubble response time to a characteristic time scale of the flow): Bubbles with a Stokes number near unity should become clustered. Because the response time of a bubble, which can be thought of as the time required for a bubble released from rest into a fluid flow to reach a steady velocity, is size dependent it is possible that naturally occurring bubble distributions created in dynamic flows are being spatially sorted by turbulent flows. If this were occurring, the pair correlation would be a function of both separation distance and bubble size. Because the single frequency multibeam observations used in this paper were unable to distinguish between different bubble sizes, the potential for bubble size sorting has been neglected.

This work has been focused on understanding the effects of clustering on short-range acoustic propagation. Clustering may also play an important role in long-range propagation scenarios, particularly in shallow water environments where there may be multiple surface bounces and interactions with near-surface bubble clouds. In these types of scenarios, clustering may be observed in two different regimes. The first is what has been analyzed here: clustering within individual bubble clouds, at scales that are less than the bubble cloud sizes. Clustering could also be used to describe the spacing of the bubble clouds themselves. In this intercloud spacing regime, a pair correlation function would be greater than 1 (i.e., showing statistical dependence on the positions of bubbles) at length scales up to the nominal size of a bubble cloud. This would occur even if the intracloud clustering was not present; it can be viewed as a statement that if there is a bubble at a given location it is likely that the bubble is inside of a bubble cloud. It would then follow that at ranges greater than the nominal size of a bubble cloud, the pair correlation function would be less than 1. That is, given the location of a single bubble, it is less likely to find a second bubble outside of the cloud. Note that the likelihood at large ranges is not zero: A second bubble may be part of the background bubble population or a second cloud. Given the form of this intercloud pair correlation function, it would be possible to analyze the effect on the acoustic field using the same framework established here. In addition to the change in correlation function, the main change would be the replacement of the free-space Green's function with a Green's function more appropriate for the case at hand (e.g., a shallow water waveguide). Thus, it is possible that at long ranges, clustering will have analogous effects to those analyzed here at short ranges.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research Ocean Acoustics program (Code 321). The methodology for using high frequency multibeam measurements to estimate pair correlation functions was developed as part of the author's Ph.D. dissertation while under the guidance of Anthony P. Lyons and David L. Bradley at the Applied Research Laboratory at Penn State. Both Lyons and Bradley also provided numerous helpful suggestions in the preparation of this manuscript, for which the author is very grateful. The author would also like to express his gratitude to Nicholas Makris at the Massachusetts Institute of Technology for extending the invitation to join the cruise on which these data were collected, to Chris Malzone and Reson for making the multibeam sonar available, and, last but not the least, to the Captain, crew, and scientific party onboard the R/V Hugh R. Sharp, all of whom weathered a rather uncomfortable storm while these data were being collected.

Benoit-Bird, K., and Au, W. (**2003**). "Hawaiian spinner dolphins aggregate midwater food resources through cooperative foraging," J. Acoust. Soc. Am. **114**, 2300.

Clay, C., and Medwin, H. (**1977**). *Acoustical Oceanography: Principles and Applications* (Wiley, New York).

Commander, K. W., and Prosperetti, A. (**1989**). "Linear pressure waves in bubbly liquids: Comparison between theory and experiments," J. Acoust. Soc. Am. **19**, 732–746.

Crawford, G., and Farmer, D. (**1987**). "On the spatial distributions of ocean bubbles," J. Geophys. Res. **92**, 8231–8243.

Dahl, P. (**2001**). "Bubble clouds and their transport within the surf zone as measured with a distributed array of upward-looking sonars," J. Acoust. Soc. Am. **109**, 133–142.

Eaton, J., and Fessler, J. (**1994**). "Preferential concentration of particles by turbulence," Int. J. Multiphase Flow **20**, 169–209.

Foldy, L. L. (**1945**). "The multiple scattering of waves, I. General theory of isotropic scattering by randomly distributed scatterers," Phys. Rev. **67**, 107–119.

Gerlotto, F., and Paramo, J. (**2003**), "The three-dimensional morphology and internal structure of clupeid schools as observed using vertical scanning multibeam sonar," Aquatic Living Resources **16**, 113–122.

Henyey, F. (**1999**). "Corrections to Foldy's effective medium theory for propagation in bubble clouds and other collections of very small scatterers," J. Acoust. Soc. Am. **105**, 2149–2154.

Jones, C. D. (**2003**). "Water-column measurements of hydrothermal vent flow and particulate concentration using multibeam sonar," J. Acoust. Soc. Am. **114**, 2300–2301.

Lax, M. (**1952**). "Multiple scatter of waves. II. The effective field in dense systems," Phys. Rev. **85**, 621–629.

Mayer, L., Yanchao, L., and Melvin, G. (**2002**). "3D visualization for pelagic fisheries research and assessment," ICES J. Mar. Sci. **59**, 216–255.

Mallock, A. (**1910**). "The damping of sound by frothy liquids," Proc. R. Soc. London, Ser. A **84**, 391–395.

Medwin, H., and Clay, C. (**1998**), *Fundamentals of Acoustical Oceanography*, Academic, San Diego, CA.

Mills, B. Y., and Little, A. G. (**1953**). "A high resolution aerial system of a new type," Aust. J. Phys. **6**, 272–278.

Monahan, E., and Lu, M. (**1990**). "Acoustically relevant bubble assemblages and their dependence on meteorological parameters," IEEE J. Ocean. Eng. **15**, 340–349.

National Defense Research Committee (**1946**). *Physics of Sound in the Sea, Division 6 Summary Technical Report* (National Defense Research Committee, Washington, D.C.), Vol. **8**.

Novarini, J., Keiffer, R., and Norton, G. (**1998**). "A model for variations in the range and depth dependence of the sound speed and attenuation induced by bubble clouds under wind-drive seas," IEEE J. Ocean. Eng. **23**, 423–438.

Shaw, R., Kostinski, A., and Larsen, M. (**2002**). "Towards quantifying droplet clustering in clouds," Q. J. R. Meteorol. Soc. **128**, 1043–1057.

Terrill, E., and Melville, W. (**1997**). "Sound speed measurements in the

surface-wave layer," J. Acoust. Soc. Am. **102**, 607–625.
Thorpe, S. (**1982**). "On the clouds of bubbles formed by breaking wind-waves in deep water and their role in air-sea gas transfer," Philos. Trans. R. Soc. London, Ser. A **304**, 155–210.
Trevorrow, M. (**2005**). "Volumetric multibeam sonar measurements of fish, zooplankton, and turbulence," Proceedings of the International Conference on Underwater Acoustic Measurements: Technologies & Results, Crete, Greece, 28 June–1 July **2005**.
Van Wijngaarder, L. (**1972**). "One-dimensional flow of liquids containing small gas bubbles," Annu. Rev. Fluid Mech. **2**, 369–396.
Weber, T. C., Lyons, A. P., and Bradley, D. L. (**2005**). "An estimate of the gas transfer rate from oceanic bubbles derived from multibeam sonar observations of a ship wake," J. Geophys. Res. **110**, C04005.
Weber, T. C., Lyons, A. P., and Bradley, D. L. (**2007**). "Acoustic propagation through clustered bubble clouds," IEEE J. Ocean. Eng. **32**, 513–523.
Wood, A. B. (**1930**). *A Textbook of Sound* (MacMillan, New York).
Wu, J. (**1981**). "Bubble populations and spectra in near-surface ocean: Summary and review of field measurements," J. Geophys. Res. **86**, 457–463.
Wu, J. (**1994**). "Bubbles in the near-surface ocean: Their various structures," J. Phys. Oceanogr. **24**, 1955–1965.
Wu, J. (**1998**). "Bubbles in the near surface ocean: A general description," J. Geophys. Res. **93**, 587–690.

Volumetric (tomographic) three-dimensional geoacoustic inversion in shallow water

A. Tolstoy[a)]
ATolstoy Sciences, 1538 Hampton Hill Circle, McLean, Virginia 22101

(Received 1 May 2008; revised 30 July 2008; accepted 4 August 2008)

Geoacoustic inversion is an important but difficult component for shallow water ocean acoustics. There are numerous methods in use to remotely determine bottom and geometric properties in both range-independent and range-dependent situations. While there have been some efforts to combine two-dimensional bottom estimates (range- and depth-dependent inversion slices) calculated around a receiving array into a three-dimensional (range, depth, and azimuthal) image of a region, there is *only one* approach in existence today, which attempts to determine *consistent* volumetric bottom properties using *multiple arrays and multiple sources*. This approach (geoacoustic tomography) will be discussed here (with final improvements via regularization).
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2981038]

PACS number(s): 43.30.Pc [EJS] Pages: 2793–2804

I. INTRODUCTION

Geoacoustic inversion (GI) is well recognized as an important tool in underwater acoustics where accurate knowledge of bottom properties leads to better understanding of the environment as well as to improved acoustic propagation predictions in shallow water ocean waveguides. Typically, GI methods involve multiple *single* slices through a scenario, i.e., single plane propagation along each source-to-receiver (SR) path. For the approach to be discussed here we combine single slice inversions in order to finally estimate *volumetric* (tomographic) environmental parameters noninvasively and thereby determine the properties for an entire three-dimensional (3D) region. We also now incorporate regularization [as suggested by Dosso *et al.* (1998)] to improve the smoothness and accuracy of the results.

We begin with the tomographic method to be developed here [based on matched field processing (MFP) as in Tolstoy, 1993] defined in detail along with its pros and cons. Originally this method was derived for a deep ocean to determine 3D ocean sound-speed profiles. However, it has been adapted in order to now estimate *any* environmental parameter, particularly bottom properties in shallow water.

Next, we will see a discussion of the Haro Strait experiment and the analysis of some of its single path data for the estimation of bottom properties. This test is of particular interest because it is the *only* shallow water test to date, which used multiple arrays of phones to receive multiple [low frequency, broadband (BB), and strong] source signals. Thus, the Haro Strait test offers the only data that may be invertible via the tomographic method.

We will note that this study concerns only water depths, not bottom properties such as sediment sound speeds, densities, attenuations, layer heights, etc. The author believes that these other properties can also be individually determined via this 3D method (thereby solving the multidimensional problem). However, a solution for non-water-depth parameters has not yet been demonstrated.

Finally, even though we cannot always invert certain Haro Strait SR paths as needed for the full tomographic inversion, we can *simulate* the final behavior of the method as though the SR inversions could have been performed. For the final volumetric inversion we get reasonable results for the resources used in the test (depending on the grid resolution used), particularly after regularization. Additionally, we also consider other simulated test regions with different source and receiving array distributions.

II. MFP TOMOGRAPHY METHOD

The tomographic method developed here was motivated by the deep ocean tomographic approach originated by Munk and Wunsch (1979) and by medical tomography. The basic idea for all these methods is to use a family of acoustic (or electromagnetic) samples passing throughout a volume to subsequently estimate volume properties. However, the original Munk and Wunsch method was *quite different* in its mathematical detail, e.g., they relied on *absolute* times on multiply distributed *single* receivers, and was intended to estimate deep ocean sound-speed profiles. The method proposed here is based on an approach developed using MFP, i.e., relying on *relative* times on multiple *arrays* of receivers.

The details of the approach to be considered here come from Tolstoy (1991, 1992) using *relative* fields along multiple arrays. This method was derived by first dividing (or gridding) a volume into a collection of cells (usually but not necessarily square in appearance from the plan view of the volume) where environmental parameters are assumed to be constant within each cell. A sample gridding of 100 cells for a 10×10 km^2 region is shown in Fig. 1 where sources ($*$) are distributed around the perimeter and three receiving arrays (R) are in the interior.

To mathematically derive the approach we originally considered the linear dependence of adiabatic normal mode

[a)]Electronic mail: atolstoy@ieee.org

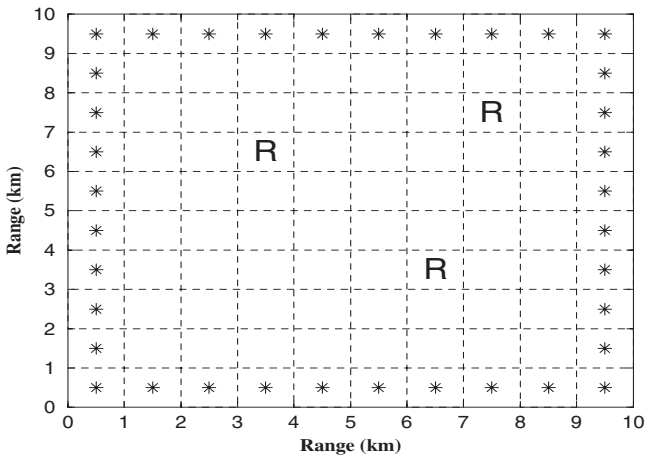


FIG. 1. Typical gridding into 100 cells for a simple square of region (10×10 km^2) with 36 sources (indicated by $^*s$) distributed around the region perimeter and three receiving arrays (indicated by Rs) in the region interior. The array placements for this region were optimized in Tolstoy, 1992.

exponents (modal phase) on sound-speed profile functions [their modified empirical orthogonal functions (mEOFs)] and on range. Assuming adiabatic normal modes (no mode coupling, see Jensen *et al.*, 1994), a simple matrix equation can be established, which when inverted gives the mEOFs [we note that if there were strong mode coupling, then the matrix equation derivation would need to be revisited (a mathematical derivation in the presence of mode coupling does not exist)]. Fortunately, adiabatic normal modes are often appropriate, particularly at the lower frequencies. See Tolstoy, 2004 for examples where this approach works well in spite of strong topographic changes. These mEOFs then result in a 3D volume of sound-speed cells. This method was demonstrated on deep water simulated data (Tolstoy, 1994) showing a highly accurate, detailed reconstruction of a highly varying synthetic ocean (Chiu and Diachok, 1992).

The GI tomographic approach for shallow water does not rederive the above deep water MFP tomography relationship. Rather, it simply assumes that the bottom parameters of each cell are constant within the cell (similar to the assumptions for the deep water mEOF coefficients) and are defined by the same kind of matrix equation as that derived originally for the deep ocean sound-speed mEOFs. In particular, for each bottom parameter p, we will assume that

$$\vec{p} = [p_m] = [\overline{p_n}\delta_n]\Delta A^{-1},$$

where $\vec{p}$ is the $1 \times M$-dimensional vector of "true" cell parameters, M is the number of cells, m is the cell index number, $\overline{p_n}$ is the "average" parameter value per nth path, n is the index number of a SR path, $\delta_n = \Sigma_m d_{nm}$ is the total distance for the nth path, d_{nm} is the distance of the nth path through the mth cell, N is the total number of SR paths, $\Delta = [d_{nm}]$ is an $n \times m$ matrix of distances, and $A \overset{def}{=} \Delta^T\Delta$ is the $m \times m$ matrix to be inverted.

Thus, the GI tomographic approach *per parameter type* involves four steps (see Tolstoy, 2000, 2001).

(1) *Estimating* for each SR path (total of N paths) the *average* per path environmental parameter(s) $\overline{p_n}$ by some inversion method where n is the SR path index.
(2) *Gridding* the region into M cells. The number of cells selected (M) is a trade-off between desirable high resolution for the region versus available acoustic sampling (N is the number of SR paths, $M \ll N$).
(3) *Inverting* the matrix A as follows:

$$A \overset{def}{=} \Delta^T\Delta,$$

$$\Delta = [d_{nm}],$$

where d_{nm} is the distance of the nth path through the mth cell.
(4) *Computing* the M-dimensional vector of true cell parameters $\vec{p}$ as follows:

$$\vec{p} = [p_m] = [\overline{p_n}\delta_n]\Delta A^{-1},$$

where $\delta_n = \Sigma_m d_{nm}$ is the total distance for the nth path.

Additionally, step 3, i.e., the inversion of the matrix A, can be improved (as recently suggested by Dosso in 2007) by the consideration of regularization. In particular, for our (second order) regularization (Press *et al.*, 1992) we shall consider that A has been replaced with the matrix $A + \lambda H$ where $\lambda \in [0, \infty]$ is optimized per scenario and

$$H = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ -2 & 5 & -4 & 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \dots & 0 \\ . & & & & & & & & \\ . & & & & & & & & \\ . & & & & & & & & \\ 0 & \dots & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & \dots & 0 & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & \dots & 0 & 0 & 0 & 1 & -4 & 5 & -2 \\ 0 & \dots & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{pmatrix}.$$

This regularization will result in improved inversion accuracy and smoothness (as we shall see in the examples in section V).

We note that the requirement of adiabaticity (used in the derivation of the deep water method) turns out to *not* typically be necessary. This was shown in Tolstoy, 2004 where inversion for water depth was shown for a number of simulated shallow water environments featuring (1) a continental shelf, (2) a Gaussian hill, and (3) a slope plus a Gaussian depression. These scenarios sometimes had quite steep bottom slopes, which would result in significant nonadiabaticity. White noise was also considered with results shown for 1% and 5% random noise added to the "average" parameter values (noise was added by selecting r from a uniform distribution of random numbers then modifying the average path parameter p by 0.01^*r^*p or 0.05^*r^*p), resulting in inversions that were still quite acceptable (maximum error less than 5.1 m out of 240 m for 1% noise).

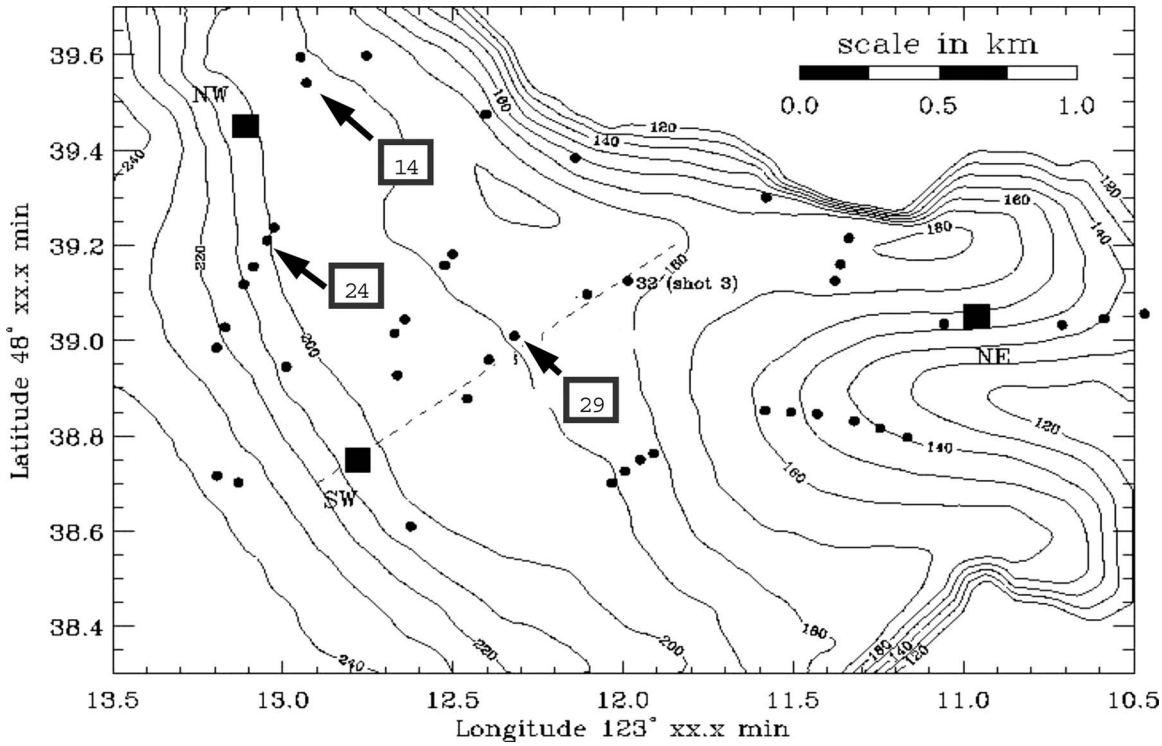


FIG. 2. The Haro Strait test region and resource distribution. The black squares show the array locations (3) while the black circles (43) show the source locations. Selected sources (14, 24, and 29) are also indicated as they were examined in detail for GI in Tolstoy, 2006a, 2006b. The nominal topography is also indicated with depths varying from 120 to 240 m. This figure is courtesy of N. R. Chapman.

Advantages of the MFP tomographic method include the following.

- It is independent of the unknown parameter (water depth, bottom sound speeds, densities, attenuations, etc.).
- It is independent of the number of parameters per grid cell.
- It is independent of the interrelationships between parameters (such as range versus water depth or sediment sound speeds versus sediment gradients).
- It is extremely fast (after individual path inversions have been performed).
- More resources (more sources, more receiving arrays, and/or phones) will usually result in improved inversion resolution.

Disadvantages to the method include the following.

- Multiple receiving arrays (preferably near vertical) are required.
- Multiple sources are required.
- Preliminary path inversions (sources to receiving arrays) are necessary to estimate average path parameters. For example, average water depth will need to be determined per path, if water depths will be determined for the region. The average parameter per path may be estimated by means of simulated annealing or genetic algorithms or by means of any method deemed most useful.
- Parameters at the source and receiving locations (their cells) must be known or predetermined, e.g., by path inversions.

We note that a critical element of the MFP volumetric (tomographic) inversion consists of preliminary SR two-dimensional (2D) slice inversions. Before a final combination of these inversions can be made, one inversion must be accomplished per SR path, and the final accuracy of the method depends on the success of the preliminary path inversions. These inversions typically rely on the receiving array design (number of phones, planned verticality, and phone spacings), the accuracy of assumed or estimated array geometry (array element localization (AEL) can be extremely helpful here), source frequencies and ranges, number of parameters estimated, interrelationships between the parameters, and more. Assuming that those preliminary inversions can and have been accomplished, the tomographic method can proceed. Unfortunately, these inversions can be problematic.

III. HARO STRAIT TEST AND EARLY ANALYSIS

The Haro Strait primer test was performed in summer 1996 in British Columbia south of Vancouver Island and deployed multiple vertical line arrays (VLAs) (3 successful VLAs, up to 16 phones each, bottom anchored, sampling rate of 1750 Hz) and multiple sources (lightbulbs, broadband =BB) at a variety of depths (nominally 30, 50, and 70 m) (Fig. 2). Usually, the signal-to-noise (S/N) ratio was very strong. The experiment was designed to measure coastal ocean processes and included GPS positioning for the source deployments plus a database of bathymetric soundings, geologic maps, and bottom core samples. One useful simplification generally confirmed by environmental sampling (see Elisseeff *et al.*, 1999) seems to be the nearly constant sound-speed profile $[c(z)=1482.5\ \mathrm{m/s}]$ present throughout the test as a result of mixing due to strong currents. This simplifica-

tion helped to reduce the dimension of the unknowns, i.e., the number of parameters to be estimated. The experiment involved the Woods Hole Oceanographic Institute, the Massachusetts Institute of Technology, the University of Victoria, and others.

This test has been the subject of much interest with regard to GI and is discussed in numerous publications and theses. In particular, Jaschke (1997) published a thesis that provides a complete description of the test and some of the first efforts made to invert a few of the individual one-dimensional (1D) acoustic data (using the NW (northwest) array and nearby sources such as shot 24 where a flat bottom could be assumed). One of the key conclusions by Jaschke (1997) was that AEL was likely to be needed as the array was very probably *not* vertical nor even a simple linear configuration, and array deformations could significantly impact GI. He was able to find some nonlinear arrays that fit the inversion data. A discussion of his (1D) effort is readily available in Jaschke and Chapman, 1999.

Next, Corré (2001) developed a dual stage method in order to increase computational efficiency and eliminate ambiguities as much as possible. In particular, Corré (2001) was concerned with the data for a single slice as well as for *two* arrays simultaneously, single as well as *two* sources at a time, and simulated as well as test data analysis. Corré (2001) also concluded that (as in Jaschke, 1997) uncertainty in array phone positions is a major problem. Results for synthetic data are shown in Corré *et al.*, 2001 and in Corré and Chapman, 2001.

Pignot and Chapman (2001) examined the SW array using three sources to determine a range-dependent (RD) 2D environment and used a RD ray code, which they developed for the inversion. Imprecise knowledge of experimental geometry, particularly for the arrays, is again acknowledged as a serious handicap for GI, and additionally, nonuniqueness is recognized to be a concern.

Most recently, Michalopoulou and Ma (2005) have examined the Haro Strait data (at the NW array) using a new estimator plus linearization applied to time domain arrivals for GI and for the estimation of geometric parameters such as source location. While the method does require prior knowledge (initialization) of water depth and phone locations to give estimates of source location, non-VLA geometries for the NW and SW arrays at particular times were also computed, which compared well with those of Jaschke (1997). Thus, their results support the likelihood that the arrays may not be linear.

After much analysis, the data are now considered to be extremely difficult to invert for geoacoustic properties as a result of *strong local currents*, which can distort and displace the *arrays* significantly as well as introduce significant uncertainty in the other geometric parameters of each data set, i.e., the ranges and depths of each source and the water depths at the sources and arrays at any given time. In addition, the bottom topography and sediment structure can be extremely RD. We see that these uncertainties (particularly AEL) are major sources of inversion difficulties for our single slices and, thus, for subsequent tomography.

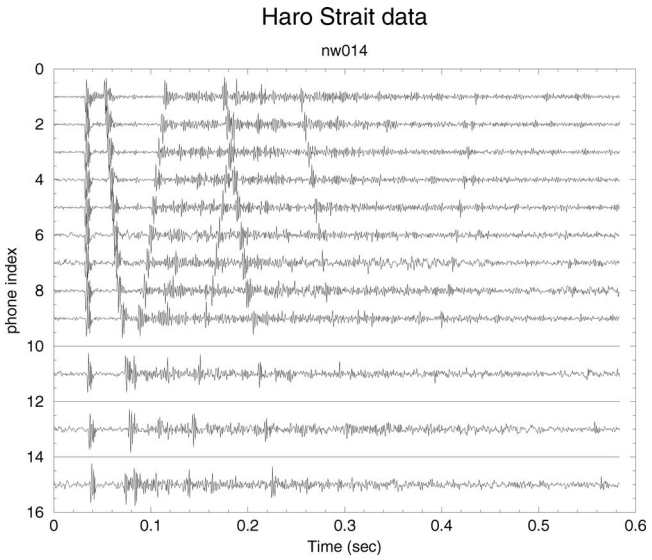


FIG. 3. NW014 single SR path data (receiving array NW and source 14) showing the data received on the indicated array phones. We note numerous strong boundary reflections.

IV. NEW SINGLE PATH HARO STRAIT INVERSIONS

As we saw in Sec. III, there have been previous Haro Strait results predicting geoacoustic parameters. These results have included estimates of sediment thicknesses, sediment sound speeds at the sediment surface, sediment sound-speed gradients, half-space sound speeds, phone ranges and depths, and source ranges and depths. Typically, some of these values have been assumed known (with nominal values) while others have been estimated. All values have been found to be within expected ranges, i.e., have been deemed acceptable. We note that only a *few* single SR paths were actually selected for inversion (not necessarily the same ones by all researchers). Questions do arise about these results. First, were some generated with too many restrictions, e.g., assuming some known parameters that were not truly "known"? How unique were the "solutions" (there are *lots* of data fits to the time domain data, see Tolstoy, 2008b)? Were the original search intervals such that *any* fit found to lie within them would be regarded as acceptable?

Recent work by Tolstoy (2006a, 2006b) has examined time domain reflection arrivals (the first four: surface reflected, bottom reflected, surface-bottom reflected, and bottom-surface reflected) where the environment can be RD by means of a linear bottom slope. For the NW array receiving source 14 the range is approximately 250 m (short range), and the data show very good separation between these first reflection arrivals (except for occasional intersections when arrivals overlap). There is also a very good S/N (a strong signal). In Fig. 3 we see the data in the time domain as seen on the array phones (the amplitudes are normalized for each phone.

Examining the time domain data it was possible (see Tolstoy, 2006a, 2006b) to derive a family of source depths, water depths at the source, water depths at the array, phone ranges, and phone depths (array restrictions are applied to disallow phones not at an appropriate distance to others or with too much snaking) such that model predictions (made

by the RD BB model RAMGEO, see Collins, 1994 and Jensen *et al.*, 1994) provided an excellent match to the data, at least for the first four (4) reflections. The details of those computations can be found in Tolstoy, 2006b. Unfortunately, the family of data fits (based on geometry only) is *excessively large* with hundreds of thousands of values, particularly if many of the phones are used, e.g., all 12 available for this path. The exhaustive set of all possible ranges and depths for the top array phone for this SR path can be seen in Tolstoy, 2008b, 2008c, where we note that *nonuniqueness* is a serious problem when matching the time domain data. We need to do more to restrict the search zone. Considering additional reflections can help to reduce the search space (Tolstoy, 2008a) but does not eliminate these concerns.

We also note that if we knew the depths of the phones, i.e., if AEL had been available for these data, then we could have estimated the source depth z_{sou} and water depth D by examining the phone depths z_{phi} at which reflection *intersections* occur. In particular, SB and BS (the third and fourth reflections: surface-bottom and bottom-surface) will intersect at $z_{\text{phi}}=z_{\text{sou}}$. This is readily deduced (using the notation of Tolstoy, 2006b) where

$$\text{dist}_3 = d_5 + d_6 + d_7 = \sqrt{r^2 + (X+y)^2},$$

$$\text{dist}_4 = d_8 + d_9 + d_{10} = \sqrt{r^2 + (x+Y)^2},$$

so that

$$\text{dist}_3 = \text{dist}_4$$

if and only if

$$X + y = x + Y,$$

if and only if

$$D_1 + z_{\text{sou}} + D_2 - z_{\text{phi}} = D_1 - z_{\text{sou}} + D_2 + z_{\text{phi}},$$

and if and only if

$$z_{\text{sou}} = z_{\text{phi}}.$$

Similarly, if we assume a flat bottom, we can find that S and B (the first and second reflections: surface only and bottom only) will intersect when

$$z_{\text{phi}} = z_{\text{sou}} + D.$$

Intersections for the later arrivals—when visible as for longer range sources—can similarly yield z_{sou} and D. That is, SBS and BSB intersect at the same depth as S and B while SBSB and BSBS intersect at the same depth as SB and BS. This intersection information plus AEL can significantly reduce the search space to determine z_{sou} and D as well as reduce the search space for more accurate phone locations.

We conclude that for the Haro Strait data, even for single SR paths, it may be *impossible* to successfully invert these data without knowledge of the array geometry, i.e., without AEL. Without AEL the search space must include a wide range of phone depths as well as source depth and water depth and is simply enormous even at one frequency. A complete and exhaustive search is not possible, and to date no solution that fits the data (at four neighboring frequencies with high MFP values ≥ 0.9) has been found for NW014.

We note that parameters only have to be constant in a grid square. For the examples to be given here, we will assume that many parameters, such as bottom sound speeds, densities, attenuations, and sediment depths, are constant throughout the region. However, for the tomographic method it really does not matter if parameters are constant throughout the region or not. We only need to preapply some inversion technique per path to estimate the average values of all parameters per path. Thus, all the average parameters per path will be determined. We can then apply the tomographic method for each parameter in turn to get the 3D picture.

V. TOMOGRAPHIC INVERSION RESULTS

Although individual SR paths may have difficulties being inverted (as seen in Sec. IV) and inversions are required prior to the application MFP tomography, we can *still* examine the volumetric method behavior assuming the inversions could have been done, i.e., via *simulated* preliminary inversions. That is, we can examine volumetric behavior assuming that average SR parameters and the geometries of the SR paths had been determined for each individual SR path.

Let us first grid the Haro Strait region into 100×100 m^2 cells (300 cells) and consider the Haro Strait bathymetry (originally shown in Fig. 2) as seen now in Fig. 4. Each cell is considered to have a constant water depth, and this gridding is considered to be rather fine scale with less than ten acoustic paths intersecting most of the cells. We have 148 cell depths to be determined. Cells with no acoustic sampling are *not* considered—there are 120 of them here, and 32 cells have sources and/or arrays contained within them for which we assume depth values known by previous inversions or measurements. This gives a total of 300 cells for the region.

Next, we consider the inversion as discussed in Sec. II, i.e., the inversion of $A \overset{def}{=} \Delta^T \Delta$. Unfortunately, the inversion of this matrix for this fine gridding will not be worth much since its condition number Λ is excessively large, i.e., $\Lambda = 2.6 \times 10^{10}$ (see Press *et al.*, 1992 for inversion computations and condition number discussions). We note that the condition number will be affected by such factors as the placement of the sources and receivers throughout the gridded region, the number of sources and receivers, and the gridding of the region (number, size, and shape). For a matrix inversion to be credible, Λ should be small where the minimum value would be 1.0 for the identity matrix.

Let us next eliminate cells which have five or fewer paths intersecting them. Thus, we hope to improve Λ by eliminating cells with little acoustic information. While this does improve Λ somewhat (it is now 4.5×10^9), this is still not a very good value. What to do? If we could, the inversion would be improved by the addition of more arrays, and/or more sources, and/or the redistribution of resources into some optimal placement, or we could decrease the resolution of the cells, i.e., increase the cell size to 200×200 m^2 lead-

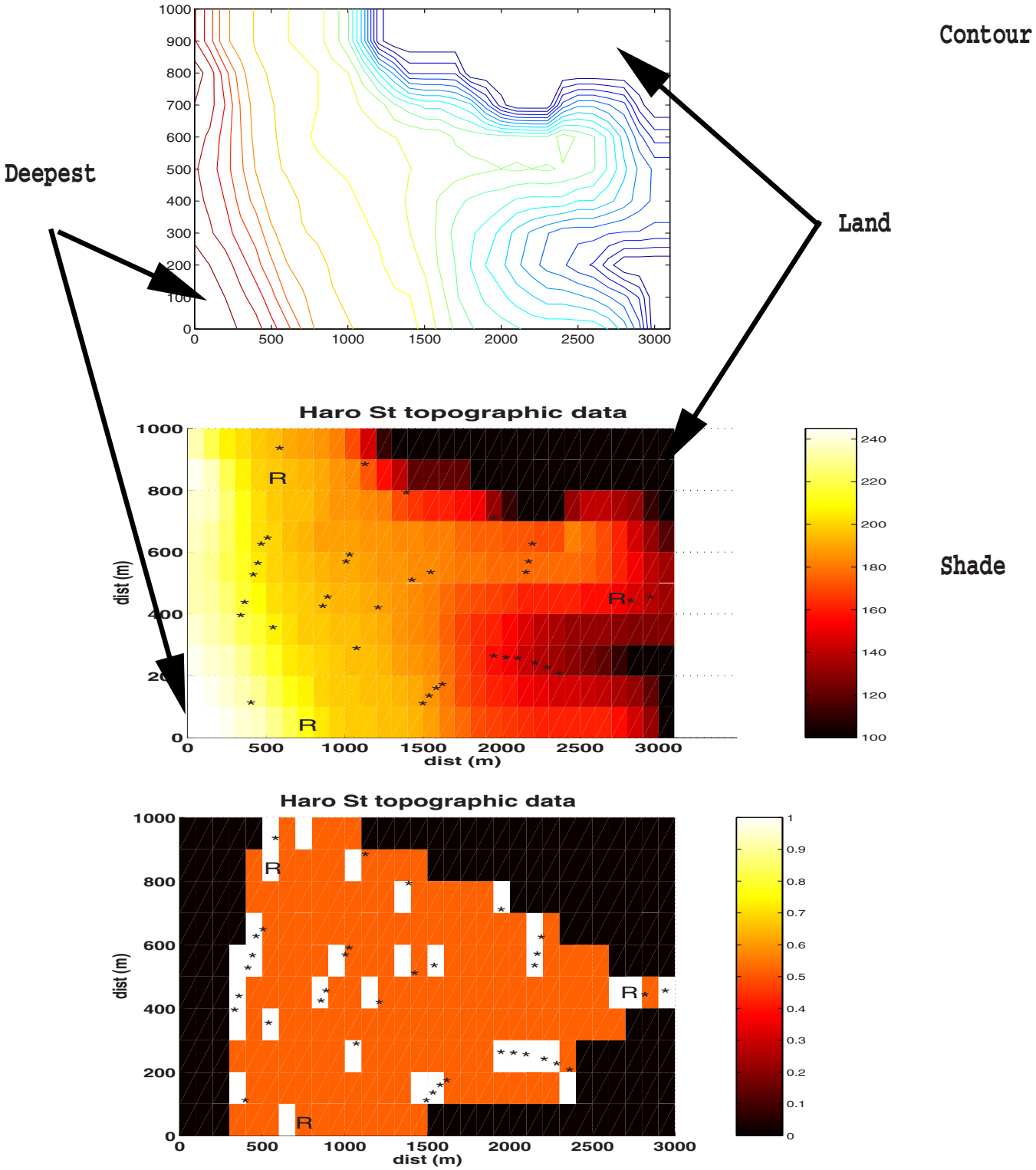


FIG. 4. (Color online) Top subplot shows the true contour water depths (as in Fig. 1). The middle subplot shows the region finely gridded into 100 $\times 100$ m^2 cells each with constant water depth indicated by the shades (scale to the right). The bottom subplot shows the region cells to be determined (148 of them), i.e., the middle shade. Cells that contain a source or an array (R) are considered to have already known depths (32 of them, white); cells outside of the SR paths (unsampled acoustically, 120 of them, black) will not be considered.

ing to 75 cells for the gridding of the region. While we cannot change the resources of the test, we can change the cell gridding.

Decreasing the cell resolution to 200×200 m^2 (shown in Fig. 5) results in fewer cells to be inverted (26 to be solved for, 26 cells with sources and/or arrays, and 23 cells with no sampling for a total of 75 cells for the region) and in improved $\Lambda \approx 8400$. After this inversion we get a rms error of 31.5 m and a maximum error of 104.7 m. This is still not a good result for a region with depths varying from 100 to

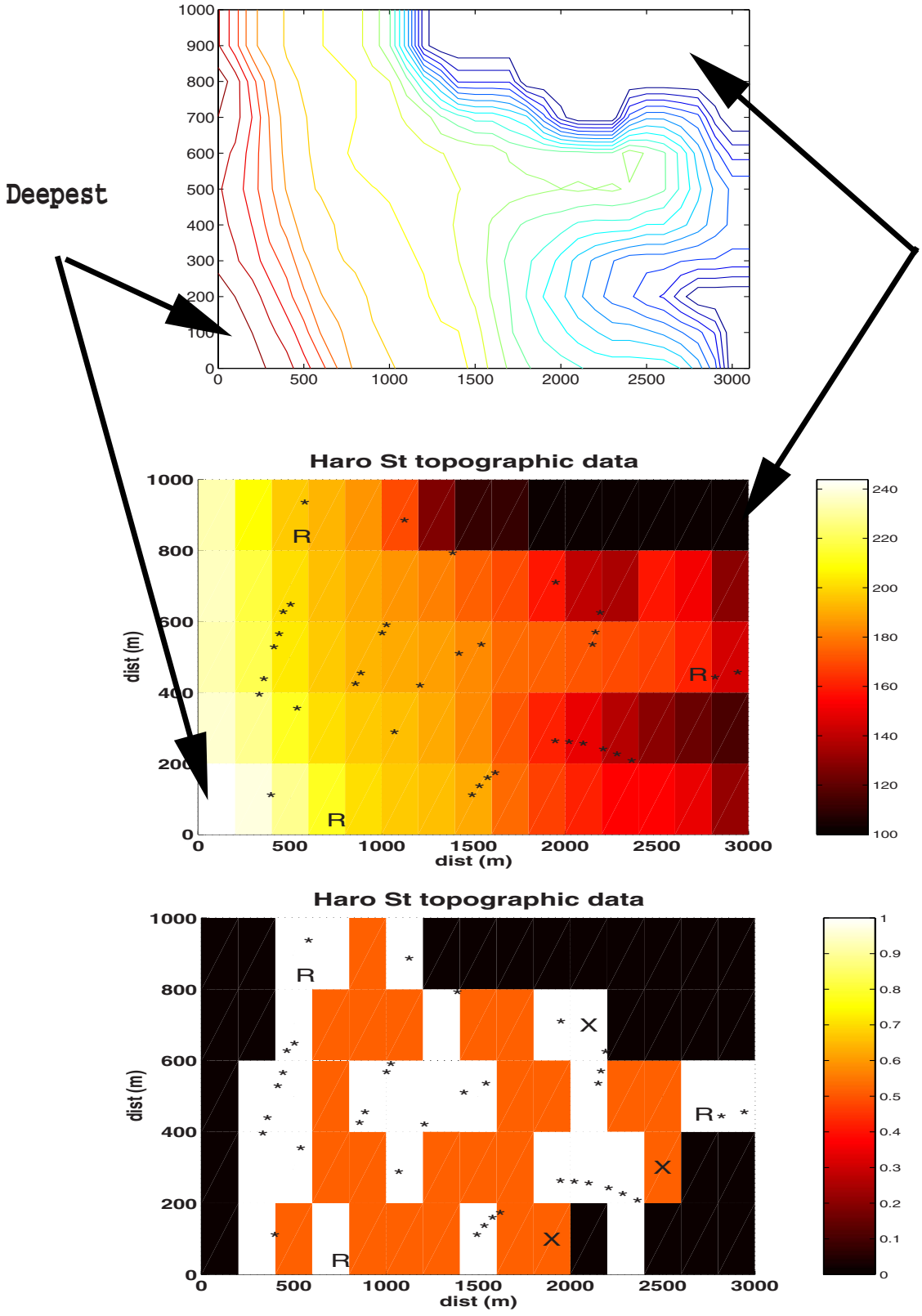


FIG. 5. (Color online) Top subplot shows the water depths (as in Fig. 1) as a contour plot of depth (region has also been gridded) with values as indicated by scale. The middle subplot shows the region gridded into 200×200 m^2 cells with the sources and array locations shown as well as the water depths as shades (scale to the right). The bottom subplot shows the *final* region cells [23 of them after 3 were eliminated (indicated by X's) for too few paths] to be determined, i.e., the middle shade. Cells that contain a source or an array (R) (or were eliminated as having too few paths) are considered to have already known depths (26 of them, white); cells outside of the SR paths (unsampled acoustically, 26 of them, black) will not be considered.

240 m. Let us again eliminate cells with few acoustic paths, i.e., with less than three paths (indicated by X's in the bottom subplot of Fig. 5). This leaves us with 23 cells to be determined and 26 cells to be ignored. We now show a rms error of 13.4 m and a maximum error of 41.3 m, with the results of the inversion shown in Fig. 6. We note that the essential

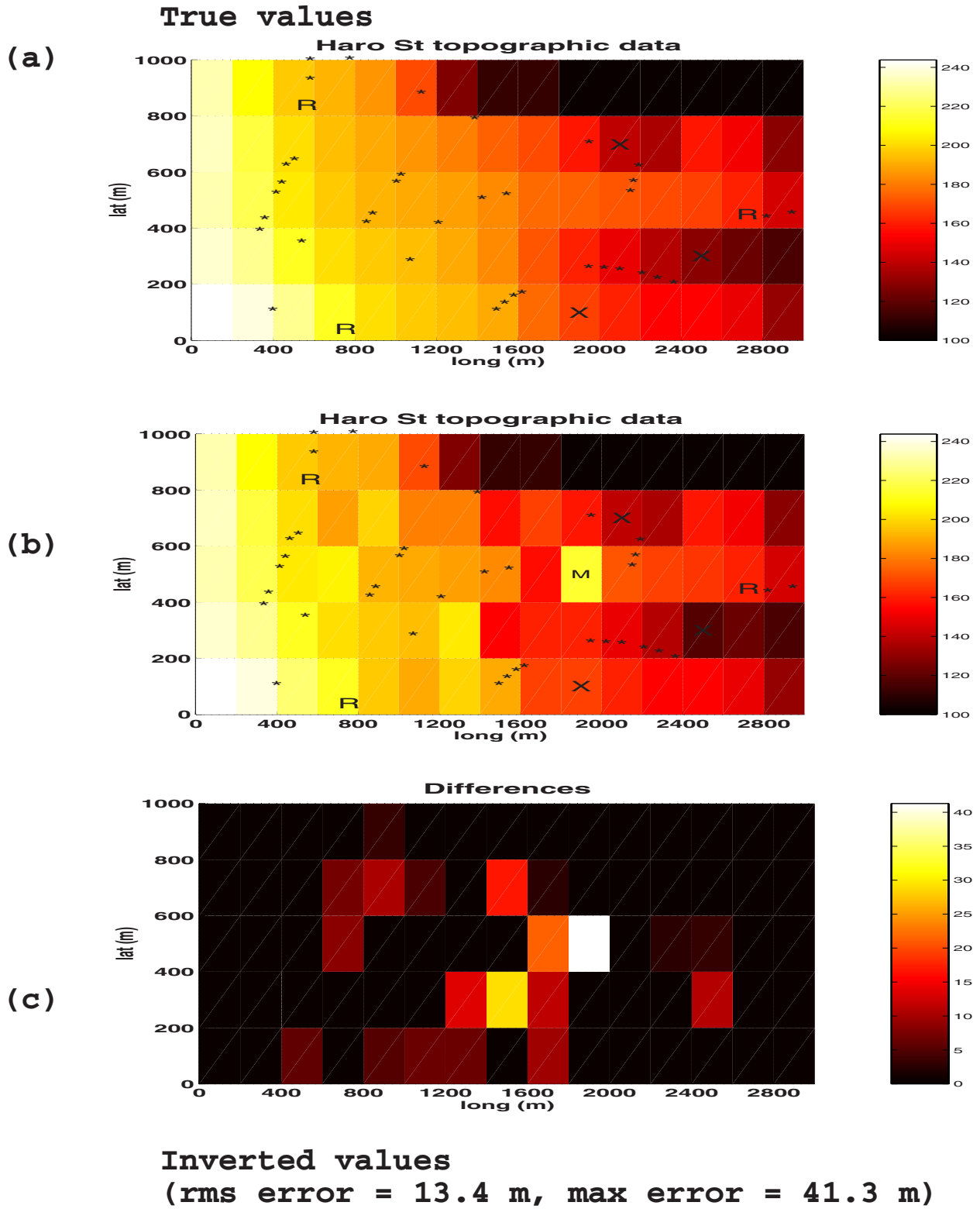


FIG. 6. (Color online) Plot showing tomographic inversion for a crudely gridded Haro Strait test where cells with three or fewer paths sampling them have been eliminated (indicated by X's) from the inversion (true depth values have been used in the bottom figure whenever no information was available). We note in (b) and (c) that the largest error in depth (41.3 m) occurs at long=1800 m, lat=400 m, indicated by M. Thus, errors can be as large as 20%. (a) shows the true values for water depth D, (b) shows the values obtained after inversion, and (c) shows the absolute value of the differences between (a) and (b).

features of the topography have been captured although the errors can be almost 20%. We note also that the largest error occurs farthest from the arrays and sources where there are the fewest intersecting paths.

Regularization for this inversion (see Dosso *et al.*, 1998) was implemented as discussed in Sec. II, and was found to improve the results. This improvement was to be expected as regularization allows *a priori* information (such as values at

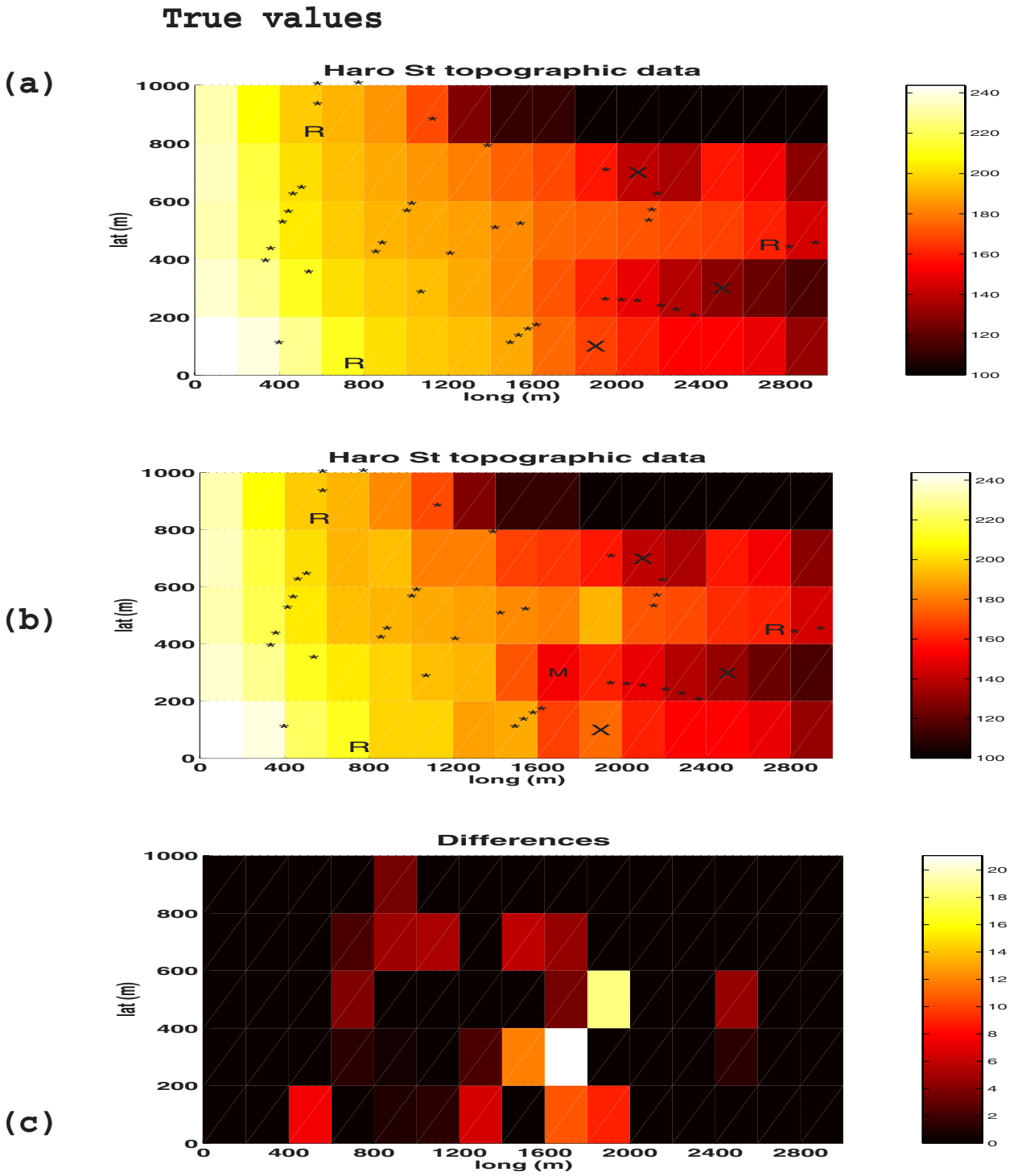


FIG. 7. (Color online) Plot showing tomographic inversion plus second order regularization for a crudely gridded Haro Strait test where cells with three or fewer paths sampling them have been eliminated from the inversion (true depth values have been used in the bottom figure whenever no information was available). We note in (b) and (c) that the largest depth error (21.1 m) is now at long=1600 m, lat=200 m, indicated by M. This error is less than 10%. (a) shows the true values for water depth D, (b) shows the values obtained after inversion, and (c) shows the absolute value of the differences between (a) and (b).

the sources and arrays) to be more influential and allows neighboring values to directly impact on the results. Thus, abrupt transitions in values between cells can be minimized. In Fig. 7 we see the improved results where now we have a rms error of 7.7 m (rather than 13.4 m) and a maximum error of 21.1 m (rather than 41.3 m) assuming the "optimal" value for $\lambda=0.01$. We now have errors less than 10%, and we determined the optimal value as that which minimized the

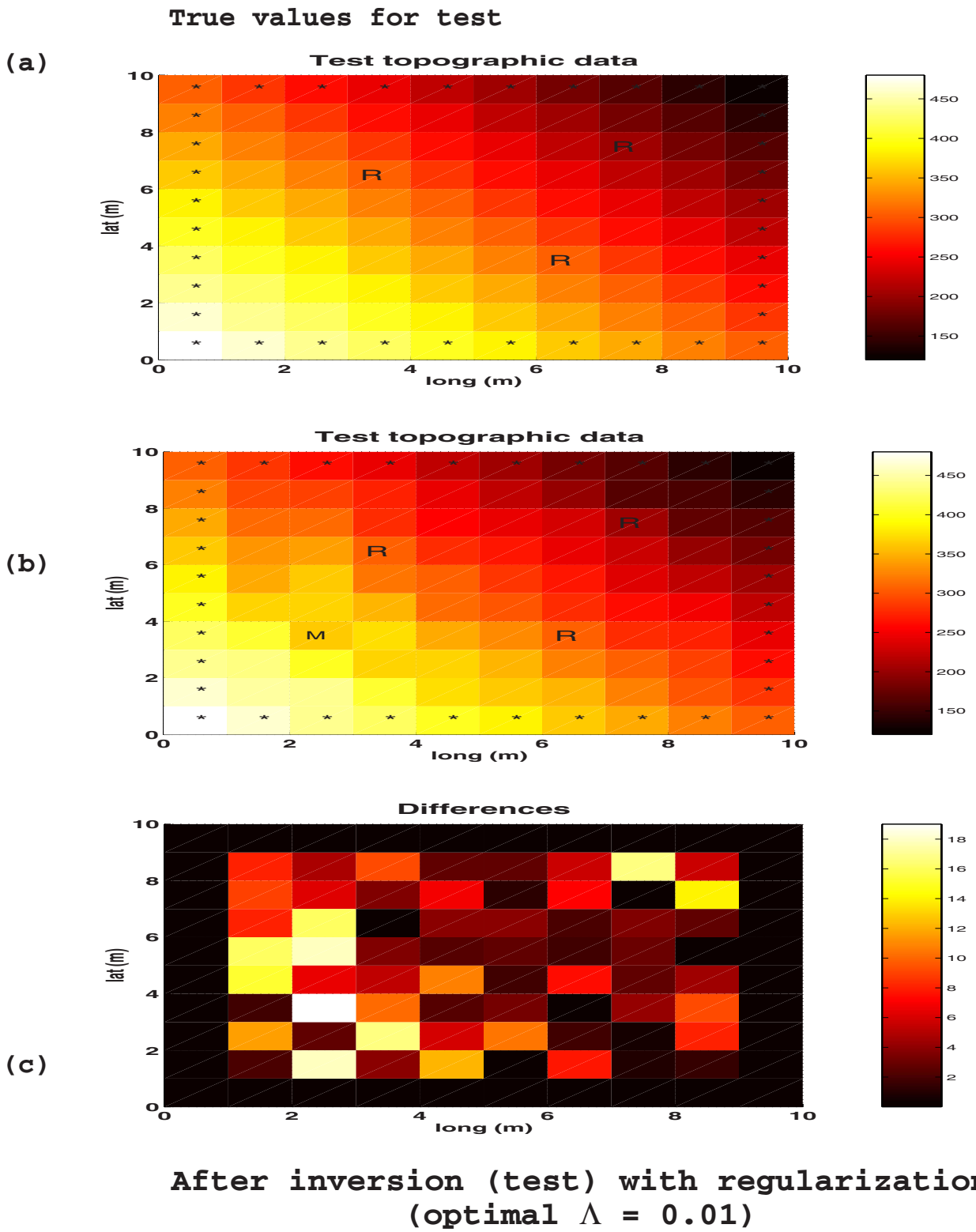


FIG. 8. (Color online) Plot showing tomographic inversion for a test case where the sources have been distributed around the perimeter and the arrays are within the region. The bottom slopes from 120 to 480 m in depth over a region 10×10 km^2 in area. We note in (b) and (c) that the largest depth error (19.8 m) is now at long=2 km, lat=3 km, indicated by M. This error is less than 5%. (a) shows the true values for water depth D, (b) shows the values obtained after inversion, and (c) shows the absolute value of the differences between (a) and (b).

inversion errors. How did we determine this optimal value for λ without knowledge of the errors, i.e., of the true parameter values? It seems to be scenario dependent, and thus can be estimated via preliminary simulation of the region and resources. Consider another test case.

Let us examine the simulation case illustrated in Fig. 8 where the water depths vary from 120 to 480 m in a well behaved manner [as seen in Fig. 8(a)]. For this case, we find that the tomographic inversion results seen in Fig. 8(b) with errors indicated in Fig. 8(c) show a rms depth error of 8.2 m

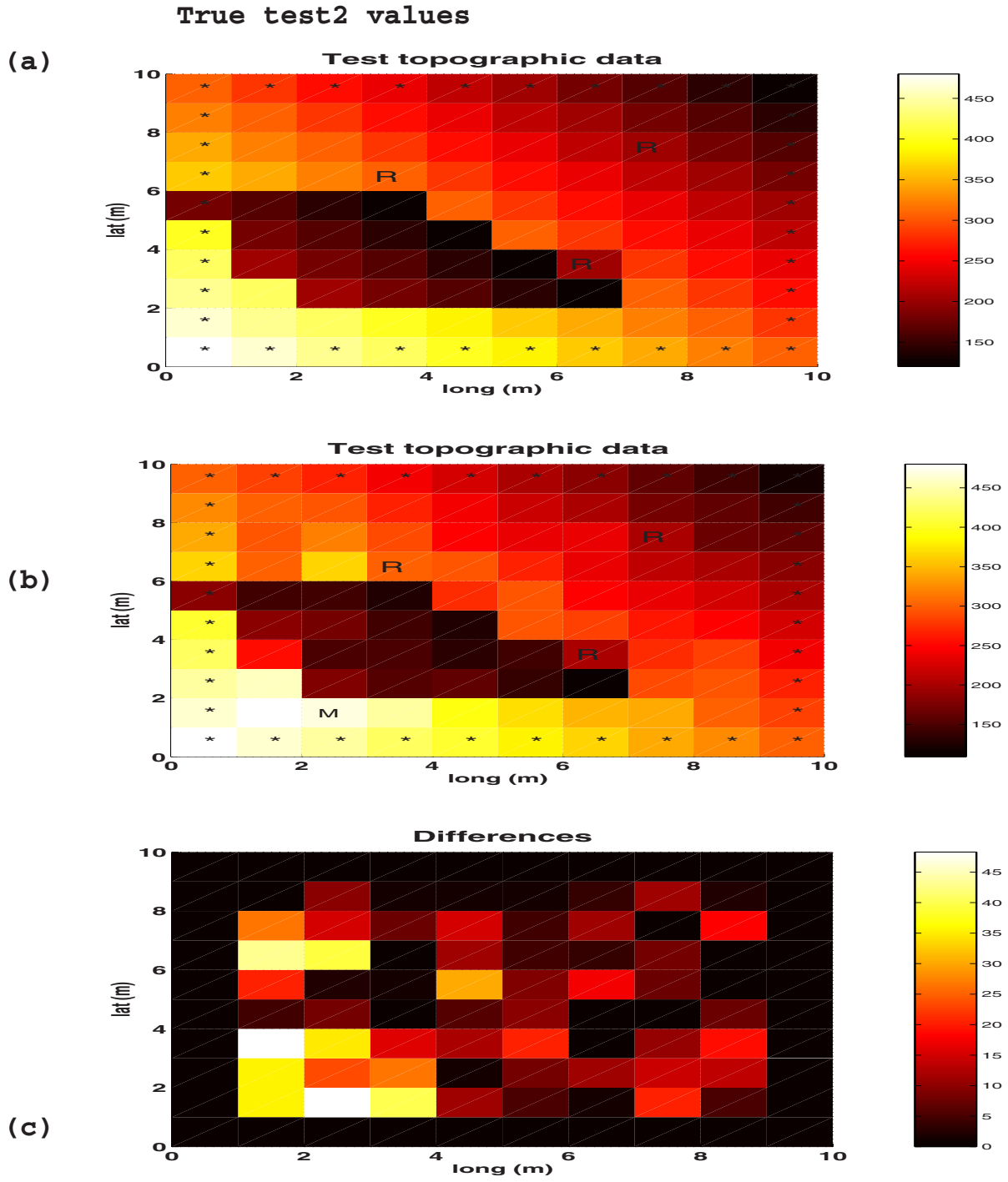


FIG. 9. (Color online) Plot showing tomographic inversion for the second test case where the sources have been distributed around the perimeter and the arrays are within the region as in the previous example. However, the bottom shows significant structure in addition to sloping from 120 to 480 m in depth over a region 10×10 km^2 in area. We note in (b) and (c) that the largest depth error (48.3 m) is now at long=2 km, lat=1 km, indicated by M. This error is approximately 10%. (a) shows the true values for water depth D, (b) shows the values obtained after inversion, and (c) shows the absolute value of the differences between (a) and (b).

and a maximum depth error of 19.8 m (assuming the optimal value of $\lambda=0.01$). Thus, the depth errors are less than 5%. If we change the topography of the simulation to that shown in Fig. 9(a) then after inversion of this more difficult region [Fig. 9(b) with errors shown in Fig. 9(c)] we have a rms depth error of 18.4 m and a maximum depth error of 48.3 m. Thus, for this topography, the errors are at most 10%. The optimal value of λ remains 0.01, and we conclude that we can determine the optimal λ for a given scenario by simulating that test independent of the parameter behavior.

VI. CONCLUSIONS

This paper has discussed the volumetric (tomographic) 3D GI approach, its mathematical basis, and its application to simulated Haro Strait data as well as to another simulated test case. We have also included second order regularization in the inversion. We know that for this method all individual SR inversions are necessary before the full 3D region inversion can be performed. However, for the Haro Strait test (the only shallow water test that has occurred with multiple distributed arrays and sources), this has *not* been possible. Incomplete array knowledge, i.e., lack of AEL, is a major drawback for Haro Strait GI leading to enormous search spaces. Moreover, nonuniqueness is also a critical issue for these path inversions. However, *simulated* tomographic GI (with and without regularization) has been performed for depths of the Haro Strait region (as well as for another simulated scenario) resulting in very respectable 3D inversions.

ACKNOWLEDGMENTS

The author would like to thank ONR for support through Grant No. N00014-06-C-0017 (and others throughout a number of years) and for direction on the Haro Strait data analysis, N.R. Chapman for his continued time and efforts to supply and discuss the data from these unique but difficult test data, E. Michalopoulou for her very helpful discussions and thoughts, and S. Dosso for his suggestions on regularization for inversion improvements.

Chiu, C. S., and Diachok, O. I. (**1992**). "Acoustical oceanography: Comparison of ocean tomography methods," J. Acoust. Soc. Am. **92**, 2324A.

Collins, M. D. (**1994**). "Generalization of the split-step pade solution," J. Acoust. Soc. Am. **96**, 382–385.

Corré, V. (**2001**). "A two-stage matched field tomography method for estimation of geoacoustic properties," Ph.D. thesis, University of Victoria, Canada.

Corré, V., and Chapman, N. R. (**2001**). "Vertical-slice matched field tomography," Acust. Acta Acust. **87**, 637–646.

Corré, V., Chapman, N. R., and Wilmut, M. J. (**2001**). "Matched field tomographic inversion to determine range dependent geoacoustic properties," J. Comput. Acoust. **8**, 295–306.

Dosso, S. E., Fallat, M. R., Sotirin, B. J., and Newton, J. L. (**1998**). "Array element localization for horizontal arrays via Occam's inversion," J. Acoust. Soc. Am. **104**, 846–859.

Elisseeff, P., Schmidt, H., Johnson, M., Herold, D., Chapman, N. R., and McDonald, M. M. (**1999**). "Acoustic tomography of a coastal front in Haro Strait, British Columbia," J. Acoust. Soc. Am. **106**, 169–184.

Jaschke, L. (**1997**). "Geophysical inversion by the freeze bath method with an application to geoacoustic ocean bottom parameter estimation," Ph.D. thesis, University of Victoria, Canada.

Jaschke, L., and Chapman, N. R. (**1999**). "Matched field inversion of broadband data using the freeze bath method," J. Acoust. Soc. Am. **106**, 1838–1851.

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**1994**). *Computational Ocean Acoustics* (AIP, New York).

Michalopoulou, Z.-H., and Ma, X. (**2005**). "Source localization in the Haro Strait using arrival time estimation and linearization," J. Acoust. Soc. Am. **118**, 2924–2933.

Munk, W. H., and Wunsch, C. (**1979**). "Ocean acoustic tomography: A scheme for large scale monitoring," Deep-Sea Res., Part A **26**, 123–161.

Pot, P., and Chapman, N. R. (**2001**). "Tomographic inversion in a range-dependent shallow water environment," J. Acoust. Soc. Am. **110**, 1338–1348.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vettering, W. T. (**1992**). *Numerical Recipes in C* (Cambridge University Press, Cambridge, MA).

Tolstoy, A. (**1992**). "Linearization of the matched field processing approach to acoustic tomography," J. Acoust. Soc. Am. **91**, 781–787.

Tolstoy, A. (**1993**). *Matched Field Processing in Underwater Acoustics* (World Scientific, Singapore).

Tolstoy, A. (**1994**). "Performance of acoustic tomography via matched field processing," J. Comput. Acoust. **2**, 1–10.

Tolstoy, A. (**2000**). "Tomographic inversion for geoacoustic parameters in shallow water," J. Comput. Acoust. **8**, 285–293.

Tolstoy, A. (**2001**). "Tomographic inversion on multiple receivers/arrays from multiple sources for the estimation of shallow water bottom properties" in *Inverse Problems in Underwater Acoustics*, edited by M. I. Taroudakis and G. N. Makrakis (Springer, New York), pp. 37–46.

Tolstoy, A. (**2004**). "Simulated tomographic geoacoustic inversion," in *Theoretical and Computational Acoustics 2003*, edited by A. Tolstoy, Y.-C. Teng, and E. C.Shang (World Scientific, Singapore), pp. 403–413.

Tolstoy, A. (**2006a**). "Haro Strait geometry (flat bottom)," J. Acoust. Soc. Am. **119**, 1388–1395.

Tolstoy, A. (**2006b**). "Haro Strait geometry (sloping bottom)," J. Acoust. Soc. Am. **120**, 1335–1346.

Tolstoy, A. (**2008a**). "Does a depth variable sound-speed profile matter for SW06 geoacoustic inversion?" J. Acoust. Soc. Am. **123**, 3106(A)(A).

Tolstoy, A. (**2008b**). "Data fit or solution?" in *Theoretical and Computational Acoustics 2008*, edited by M. Taroudakis and P. Papadakis (University of Crete).

Tolstoy, A. (**2008c**), "Volumetric (tomographic) 3-D geoacoustic inversion," in *Important Elements in: Geoacoustic Inversion, Signal Processing, and Reverberation in Underwater Acoustics*, edited by A. Tolstoy (Research Signpost, India).

Tolstoy, A., Diachok, O., and Frazer, L. N. (**1991**). "Acoustic tomography via matched field processing," J. Acoust. Soc. Am. **89**, 1119–1127.

Comparison of measured and modeled temporal coherence of sound near 75 Hz and 3683 km in the Pacific Ocean

John L. Spiesberger
Department of Earth and Environmental Science, 240 S. 33rd Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6316

(Received 10 August 2007; revised 24 July 2008; accepted 28 July 2008)

The hypothesis to test is internal gravity waves following a Garrett–Munk spectrum are sufficient to explain temporal coherence of sound at 3683 km in the Pacific Ocean for a signal at 75 Hz and a pulse resolution of 0.03 s. Signals from a 20 min transmission are collected on a towed array. After correcting the data for what likely appears to be acceleration of the receiver, the probability distribution for multipath coherence time is very similar to that obtained from Monte Carlo simulations of the impulse response. The most likely coherence time is 20 min, the longest that can be measured with a 20 min transmission. Predictions of multipath temporal coherence and amplitude fluctuations appear accurate enough to make useful predictions of channel capacity.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977676]

PACS number(s): 43.30.Re [WC] Pages: 2805–2811

I. INTRODUCTION

Temporal coherence of low-frequency sound is important for detection, location, acoustic communication, acoustical oceanography, and theoretical studies. We estimate coherence with a passive towed research array. Signals originate from an acoustic source (75 Hz and 0.03 s resolution) at Kauai[1] at 3682.5 km distance (Fig. 1). We test the hypothesis that coherence can be predicted from fluctuations of internal gravity waves obeying the Garrett–Munk spectrum.[2,3] Standard methods in acoustics and oceanography are used with no effort to fit models to data. A Monte Carlo technique yields time-varying impulse responses at the receiver by evolving the internal wave field with a linear dispersion relation. We adopt a minimalist approach toward the understanding causes of coherence. Unless a strong inconsistency is found between the model and data, there is not a strong reason to include additional phenomenon or modify model parameters. In this and previous works,[4,5] parameters are taken from original literature without modification. If temporal scales from this experiment agree with the model, it would be the model's third success[4,5] out of four attempts. The fourth attempt was inconclusive because the experimental analysis by others appears to be not quite complete.[6] Using different models in another experiment, comparison of temporal coherence of acoustical mode one was confounded by difficulty in removing motion of instruments.[7] However, when instruments do not move, or when their role can be well enough understood, Monte Carlo methods appear to yield accurate estimates of coherence. Particularly important may be applications in communication, where adaptive methods are commonly used to suppress effects from instrumental motion.

II. EXPERIMENT AND DATA PROCESSING

Data come from an acoustic source at Kauai[1] and a towed array in the Gulf of Alaska (Fig. 1). The source is placed at the bottom at 816 m depth at 22.349 °N, 200.43033 °E. The power, timing, and transmission characteristics are controlled by connecting the source by cable to a computer on land. Its time base is linked to an atomic standard. A broadband signal is transmitted at a center frequency of 75 Hz. Carrier phase is modulated every two cycles using a linear maximal shift register sequence having 1023 digits. First zeros in the emitted spectrum are at $75/2=37.5$ Hz on either side of 75 Hz. The pulse resolution is $1/37\sim0.027$ s.

The signal periodicity and level are $(1023)2/75=27.280$ s and 195 dB *re* 1 μPa at 1 m, respectively. At 04:00 Greenwich Mean Time on 6 September 2002, the source transmitted 44 consecutive periods lasting 20 min. See Ref. 1 for details on the transmitted code.

The receiver is a horizontal array towed at 131 m depth at 53.5219302 °N, 215.319727 °E. Location is written with greater precision than its accuracy of a few kilometers. This precision is given so others can model this section with identical coordinates. It is not possible to compare with models absolute time of signal propagation between the source and the receiver with O(1) km error.

The geodesic length is 3682.6 km from source to receiver (Fig. 1). During signal arrival, the array speed is 0.2 m/s with heading at 286 °T. Since the bearing angle of the geodesic at the receiver is 25.9 °T, the incoming signal makes an angle of 80.1° with respect to ship heading. During 20 min of data reception, the ship travels 0.24 km and is 0.041 km closer to the source at the end of 20 min.

III. MODELS

A. Environment

The section of sound speed is computed using Del Grosso's algorithm[8] and Levitus' climatological averages[9] of temperature and salinity for summer. The depth of minimum speed varies from 800 m at the source to 90 m at the receiver. Since the acoustic models use Cartesian coordinates, the sound speed profiles are translated to Cartesian coordinates using the Earth-flattening transformation.[10]

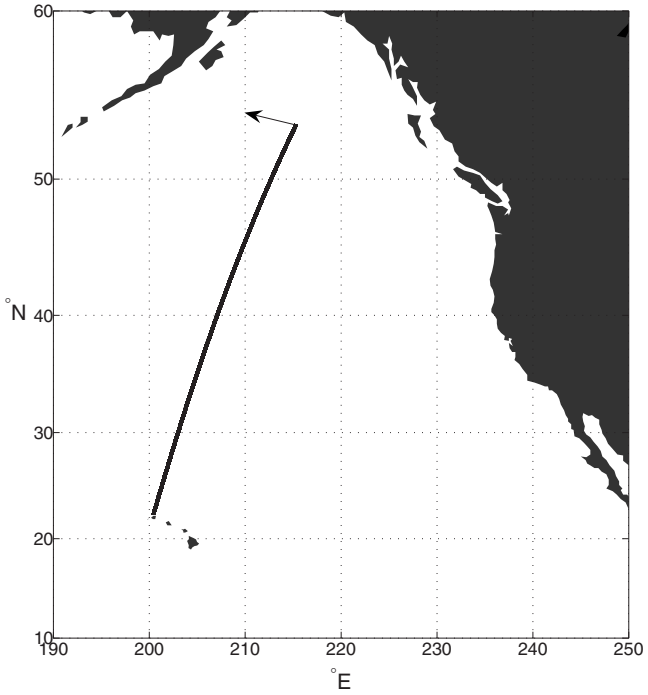


FIG. 1. Geodesic between Kauai source and towed receiver with heading shown.

Internal waves are modeled with the complete Garrett–Munk[2,3] spectrum. It differs slightly from the method by Colosi and Brown[11] who used WKB scaling to obtain internal wave modes in the vertical coordinate. We solve the equation without WKB scaling. Reference 12 completely describes the model used here.

Currents are ignored, being two orders of magnitude less than sound speed perturbations arising from adiabatic vertical displacements of water in the upper ocean. Perturbations are added to climatology of sound speed described above. The internal wave modes are precomputed and retrieved as needed at range intervals of 80 km to account for changes in water depth, buoyancy frequency, and sound speed. The vertical displacements of modes are set to zero at surface and bottom. For each 80 km interval, a three-dimensional field of internal waves is computed in a box of 80×80 km^2 by D m, where D is the average depth of ocean in that interval. A vertical slice through the box gives vertical displacements along the geodesic for any desired section. Temporal evolution of the field is governed using the linear dispersion relation. Energy of the internal wave field is taken to be that specified by Garrett and Munk.[2,3]

Bathymetry along the section consists of a steeply sloping region on the slope of Kauai, followed by an abyssal region with seamounts (Fig. 2). Within 100 km of the source, bathymetry is measured with a Seabeam survey[13] (B. Howe, personal communication). A bathymetric database is used elsewhere.[14]

The parabolic approximation of the acoustic wave equation requires parameters to define acoustic propagation in the seabottom. Those chosen (not shown) may or may not be close to those at the site. Other values would very likely not significantly change coherence times of modeled multipath. They would likely change their amplitudes, but this does not seem to be important for comparing the measured and modeled probability distributions of coherence time.

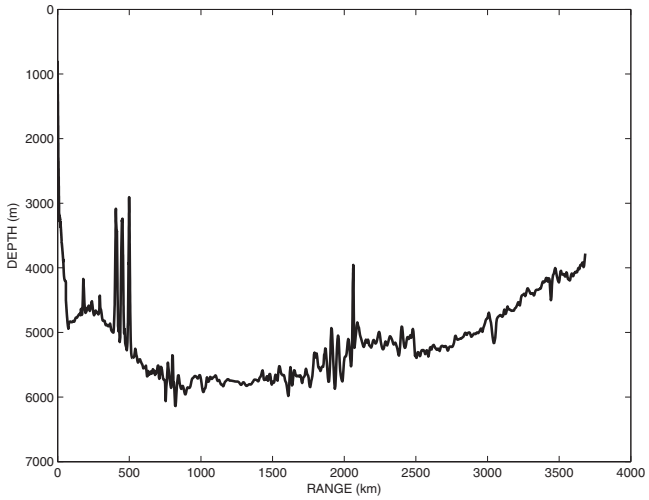


FIG. 2. Bathymetry between acoustic source at 0 km and receiver.

B. Acoustic models

Fans of rays are traced using a program, ZRAY, that is a modification of ray.[15] Eigenrays are found using another program. These programs have been used to identify acoustic paths before.[16] Rays reflect specularly from the bottom. Both geometric and nongeometric arrivals are found. Geometric types are those that pass through the source and receiver. Nongeometric types are those that provide energy at the receiver on the shadow sides of caustics. For lack of a more reliable value, rays that reflect from the bottom suffer an attenuation of 3 dB/bounce. The sound speed field used for tracing rays is identical to that used for the sound speed insensitive parabolic model[17] at its computational grid (below).

This parabolic approximation[17] computes a two-dimensional field of sound along the geodesic from 0 to 8000 m depth. Tests suggest that travel times of pulses are computed with an accuracy of a few milliseconds.[17] The impulse response is computed by applying an inverse Fourier transform to many single-frequency computations. Vertical slices of sound speed are obtained from a three-dimensional field of internal waves (Sec. III A). Numerical convergence is obtained by halving grid sizes until the answers do not change significantly. Parameters chosen lead to convergence of a normalized cross-correlation coefficient of 0.99 or greater. We use a vertical grid spacing of 1.95 m. The horizontal grid spacing varies between 33 and 150 m. A separation of 33 m is used when bathymetry is steep near source and seamounts.

IV. IMPULSE RESPONSE

Using standard techniques, a nonadaptive time-domain beamformer is used to steer a beam toward the source. The beam is much wider than any variation in signal direction. Next, each separate 27.280 s M-sequence period is adjusted for various Doppler corrections and correlated with a replica of the transmitted sequence.[18] Replica correlation with a linear maximal shift register sequence compresses 27.280 s of

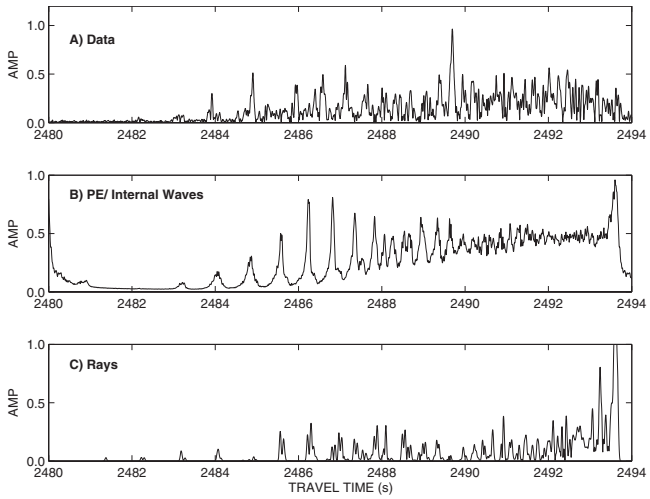


FIG. 3. Comparison between measured (a) and modeled impulse responses [(b) and (c)], each on its own arbitrary amplitude scale. Travel times of data are shifted to approximately correspond with models. The parabolic approximation model (b) is the result of incoherently averaging 150 impulse responses through an evolving field of internal waves. The ray model (c) does not incorporate fluctuations from internal waves.

energy from each acoustic path into a pulse of duration $2/75=0.02667$ s with a theoretical gain of $10\log_{10}(1023)=30$ dB.

Search in Doppler space is done to maximize the average signal-to-noise ratio of multipath due to an equivalent relative speed between the source and the receiver. Search increment is fine enough that one discerns changes of an average multipath signal-to-noise ratio of 0.1 dB or less near optimum correction. Individual fading of any single acoustic path is not caused by this procedure We obtain sufficient signal-to-noise ratio to examine the impulse response. To avoid end effects, 42 of 44 transmitted periods are processed, with the first and last unused.

Peak signal-to-noise ratio varies within the 42 periods between 20 and 25 dB (not shown). After applying an optimum Doppler correction for each period, the peak signal-to-noise ratio is 33 dB in the coherent average of these 42 periods [Fig. 3(a)]. A standard weighted average,

$$\bar{x}=\frac{\Sigma_{i=1}^{42}x_i/\sigma_i^2}{\Sigma_{i=1}^{42}1/\sigma_i^2},$$

is used to emphasize periods with higher signal-to-noise ratios, where σ_i^2 is the noise variance of the ith M-sequence period. If all 42 periods had the same signal-to-noise ratio, a $10\log_{10}42=16$ dB gain would be obtained over any individual period when the noise was uncorrelated and the signal was perfectly correlated. The gain to 33 dB in the weighted coherent average, from 20 to 25 dB/period, is a bit less than ideal. Earliest paths arrive near 2483 s and end just before 2494 s [Fig. 3(a)].

The parabolic approximation[17] is used to compute 150 realizations of impulse response at 4 min intervals through a time-evolving field of internal waves. White Gaussian noise is added to each impulse response having about the same peak signal-to-noise ratio as data. When the record-averaged peak signal-to-noise ratio is 20 dB, the incoherent average of the model has about the same time extent as measured (Fig. 3(b)]. Modeled energy between 2480 and 2482 s comes from periodicity of the impulse response obtained by Fourier transforming a finite number of frequencies.

If we had measurements of the impulse response over a day or longer, previous research indicates that likely good correspondence would be found between incoherently modeled and measured impulse responses.[16,19] However, a single measurement is not enough to show up all paths, many of which intermittently fade due to fluctuations in the sea.

The impulse response from ray theory also has about the same time extent as observed [Fig. 3(c)]. Rays are traced through climatological variations without internal waves. Amplitude of the final set of rays is truncated from 1.9 to unity near 2494 s so amplitudes of weaker arrivals can be seen more easily. Earliest arriving rays have upper turning depths varying from about 240 m near the source to 10–50 m near the receiver. Rays arriving last have upper turning depths varying from about 750 m near the source to 60–70 m near the receiver. These ray depths gradually rise as sound travels into colder waters where depth of minimum sound speed shoals. Most clumps of arriving energy are composed of temporally unresolved rays.

We did not trace rays through an internal wave field to avoid effects of ray chaos. Effects of internal waves on sound are better portrayed with the parabolic approximation[17] [Fig. 3(b)].

V. TEMPORAL CORRELATION

A. Doppler corrections

In Sec. IV, we found that all 42 impulse responses could be coherently averaged to increase signal-to-noise ratio over that for any individual impulse response after separately correcting each for Doppler. The fact that a Doppler shift is needed begs the question as to the roles of receiver motion and ocean in setting coherence time. We do not have accurate enough estimates of receiver speed to separately remove its effects. In a previous experiment with stationary source and towed array, we found coherence time to be insensitive to Doppler speed correction used during reception of various M-sequences.[5] For example, one could use the best Doppler speed correction for the first M-sequence period to correct the remaining periods. This result yielded coherence times that looked like those obtained by separately correcting each M-sequence period. This likely happened because tow speeds were quite steady. We are not so lucky in this experiment. Velocity of the receiver appears to change enough so that we cannot use a Doppler correction from the first period to correct the remaining periods without seriously affecting the estimates for coherence time (not shown). Consequently, there is less information concerning the ocean's role in limiting coherence time.

The array is towed at 0.2 m/s nearly perpendicular to direction to the source. Doppler corrections are small, starting at 0.0886 m/s for the first M-sequence period and ending at −0.0125 m/s for the last. The decrease is not monotonic but shows a definite and gradual trend toward lower speeds during 20 min of reception (not shown). The source lies on

the ocean bottom, so the first thing to check is whether these Doppler corrections are primarily due to acceleration of the receiver or fluctuations in the sea.

Change in sonic travel time due only to acceleration a of a receiver in geophysical time interval $\delta\tau$ is

$$\delta t_a = a\delta\tau^2/(2c), \tag{1}$$

where c is the signal speed, $a=\delta v/\delta\tau$, and change in Doppler speed is δv. For oceanographic causes, consider a barotropic current over distance R of form $u=u_0 \cos \omega\tau$, where u_0 is the magnitude and ω is the radian frequency. The effect of this current on travel time is $\delta t_u=-uR/c^2$ to first order. Change in travel time in interval $\delta\tau$ due to the temporal change of the current is

$$\Delta\delta t_u = u\omega R\delta\tau \sin(\omega\tau)/c^2, \tag{2}$$

with the maximum value found for $\sin \omega\tau=1$. Equating maximum value of Eq. (2) with δt_a, we find

$$\frac{u_0 R}{T} = c\delta v/4\pi, \tag{3}$$

where $T=2\pi/\omega$. For the observed Doppler changes of $\delta v=0.1$ m/s and $c\sim 1479$ m/s, the right side of Eq. (3) is 11.8 m^2/s^2. As for the left side, typical values of barotropic currents in open ocean are 0.01 m/s over scales of the basin, so take $R=3600$ km, and consider a semidiurnal period of 12 h. Then the left side is $0.01\times 3.6\times 10^6/(12\times 3600)=0.8$ m^2/s^2. This is an order of magnitude smaller than the 11.8 m^2/s^2 needed to account for a 0.1 m/s change in Doppler.

At ocean-basin scales, nonstationary oscillations in the north Pacific lead to changes in travel time between a few and tens of milliseconds at periods between 2 and 20 h.[20] These measurements appear to be the only other oceanographic candidate that might explain change in Doppler. The origin of the oscillations is unknown, but u_0R/T can be estimated since we know T and $u_0R=-c^2\delta t$. For $T=2$ h and $\delta t=0.002$ s, $u_0R/T\sim 0.6$ m^2/s^2. For longer periods, $T=20$ h and $\delta t=0.01$ s, so $u_0R/T\sim 0.3$ m^2/s^2. Both are an order of magnitude too small to account for 11.8 m^2/s^2 needed to account for observed changes in Doppler. Section V C below will explain why the GM spectrum[2] itself does not account for observed corrections in Doppler. We conclude that observed 0.1 m/s change in Doppler is probably due to acceleration of the receiver.

B. Data

With one 20 min transmission, there are many fewer degrees of freedom to estimate coherence time than a different experiment where transmissions were continuous for five days.[4] The previous experiment computed a standard normalized correlation coefficient of coherence time of the entire impulse response lasting 4 s. In this paper, degrees of freedom are increased by considering coherence of many small windows of arrival time from each of 42 impulse responses. We also use a different definition of coherence time. We might have used a traditional definition of coherence time based on a lag yielding a normalized correlation coefficient of e^{-1}. However, we see below that many paths are coherent even at the longest explorable lag of 20 min.

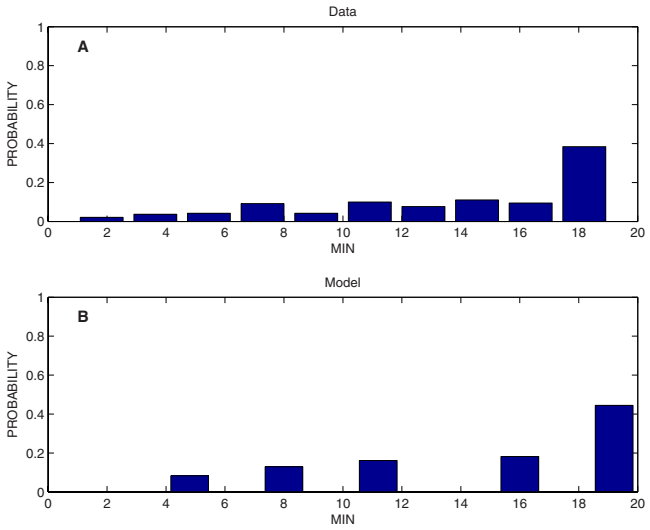


FIG. 4. (Color online) (a) Probability of temporal coherence within 380 analysis windows of 0.03 s each covering 11.4 s duration of the entire measured impulse response. Probability that any analysis window has the given coherence time is shown up to 19.06 min (42 M-sequence periods). Each M-sequence period is separately pre-processed for the best Doppler correction yielding highest average signal-to-noise ratio in the impulse response. (b) Same as (a) except computed from modeled impulse responses without any Doppler correction. Average peak signal-to-noise ratio in the model is set to 20 dB, about the same as measured.

As explained before, the data are Doppler corrected to suppress acceleration of the receiver. The resulting impulse response for each of 42 periods is subdivided into 380 windows of travel time of duration 0.03 s each, which is about the temporal resolution of the transmitted signal. Windows are chosen to cover energetic arrivals lasting 10 s. For each window, we compute the peak signal-to-noise ratio from each coherent average,

$$a(t,m) = \frac{1}{m}\sum_{i=1}^{m} r_i(t), \quad m = 1,2,3,\ldots,42, \tag{4}$$

where $r_i(t)$ is the real-valued time series at the receiver within one window, and t is the arrival time. Letting $\hat{m}$ denote the value of m yielding the maximum signal-to-noise ratio, coherence time is defined as $T=\hat{m}27.628/60$ min. For example, if the coherent average of the first eight M-sequence periods has largest signal-to-noise ratio, the coherence time is $8\times 27.628/60$ min ($m=8$). An empirical probability distribution is plotted for these 380 windows [Fig. 4(a)]. The most likely coherence time is 19 min. It occurs with a probability of 0.4. Lesser coherence times are distributed between 2 and 16 min. Coherence time of the signal may be longer than 19 min, but we cannot explore this possibility with 42 transmissions of this M-sequence.

It is sometimes thought that signal-to-noise ratio of a coherent integration increases monotonically to some peak then decreases. However, rise to maximum need not occur monotonically.[6] In this experiment, signal-to-noise ratio often does *not* increase monotonically. For example, signal-to-noise ratio decreases for the largest peak in the impulse response at about 1 min and again at 6 min of integration time

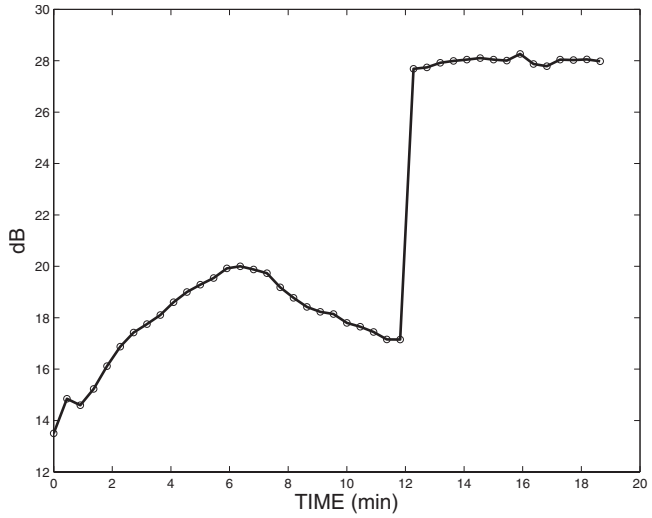


FIG. 5. Signal-to-noise ratio vs time of coherent integration. The window corresponds to the largest peak in the impulse response of the data [Fig. 3(a)]. The circles occur at intervals equal to the M-sequence period (27.280 s).

before rising to 28 dB near 12 min (Fig. 5). Considering 380 windows of arrival time, many coherence times decline before eventually rising (not shown). The model exhibits similar behavior (not shown). Temporary decreases in signal-to-noise ratio are not due to sudden increase in noise. Instead, noise varies little during any of the 42 M-sequences (Fig. 6). Neither model nor data show signal-to-noise ratio increasing monotonically with number of periods m as $10\log(m)$ dB. Our definition of coherence time does not depend on whether signal-to-noise ratio increases monotonically.

We found no significant difference in coherence time between the early and late energies. Both have coherence times with probability distributions similar to that in Fig. 4(a) (not shown). Window width is increased from 0.03 to 0.09 s to investigate if coherence time changes. Coherence time could decrease with a number of independent multipath in the window. However, there is no significant difference from that in Fig. 4(a) with a 0.09 s window (not shown). Evidently, there are not significantly more independent oscillations in a 0.09 s than 0.03 s window.

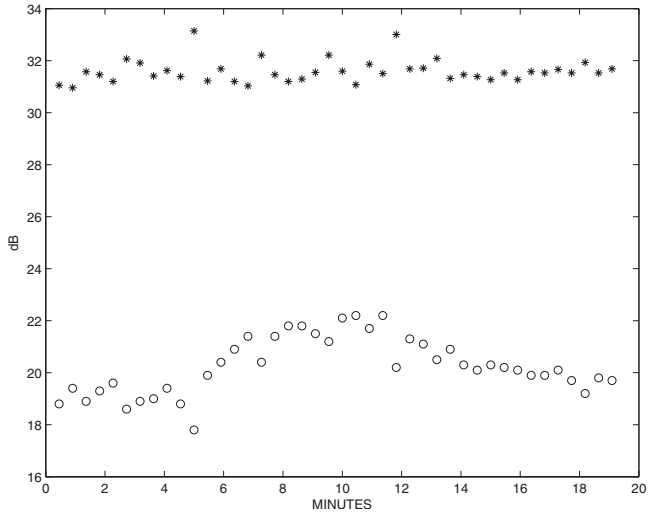


FIG. 6. Standard deviation of noise (*) and average signal-to-noise ratio of biggest 80 peaks (o) for each of 42 M-sequence periods for optimum Doppler correction. Both quantities are in dB.

C. Model

The parabolic approximation[17] yields an impulse response for 150 records at 4 min intervals. (Internal waves evolve by 4 min between computations.) Each impulse response is subdivided into $W \equiv 365$ adjacent windows of width 0.03 s. This covers the modeled impulse response lasting 10 s (Fig. 3(b)). White Gaussian noise with mean zero and variance σ^2 is added to each record. Signal-to-noise ratio of each realization is the same as the data in the following sense. Let peak amplitude of record j be $\hat{a}_j$, $j = 1, 2, 3, \ldots, 150$. The record-averaged peak amplitude is $\bar{a} = 150^{-1}\Sigma_{j=1}^{150}\hat{a}_j$. Variance is determined by solving for σ^2 in $20 = 10\log_{10}(\bar{a}^2/\sigma^2)$(dB). This ensures that record-average peak signal-to-noise ratio is the same as measured.

A bootstrap scheme is used to estimate the coherence time for each of 365 windows. First, we select at random $B = 3000$ different starting records among 150 possibilities. Direction of the coherent average is selected at random to go forward or backward in geophysical time with respect to starting record. A total of five records are added together in the randomly chosen direction. End point problems are handled by choosing a direction that would not extend below 1 or above 150. With five records, we explore coherence times up to 5 records $\times$ (4 min/record) = 20 min. For each starting record, coherence time is computed by selecting the number of records n, yielding largest signal-to-noise ratio where n can go from 1 to 5. Coherence time is $4n$ min. Letting coherence time for bootstrap b of travel time window k be T_{bk}, there are $BW = 3000 \times 150 = 450\,000$ estimates of T_{bk}. An empirical probability distribution is computed from these [Fig. 4(b)]. It resembles that from data. The most likely coherence time is 19 min, occurring with a probability of 0.4. Histogram bars have slightly different centers for model and data because model and data are available at 240 and 27.28 s intervals, respectively. The results are almost the same if the record-average signal-to-noise ratio is 25 instead of 20 dB (not shown).

The GM spectrum does not account for Doppler corrections of 0.1 m/s. To see why, we estimate the average change in acoustic travel time between each of 150 impulse responses using the phase and amplitude of each model output. The technique is described elsewhere.[20] Travel times change by less than 0.001 s between the 1st and 150th records, which is a tenth of a cycle at 75 Hz. This is too small to cause observed Doppler corrections.

Of 365 modeled windows, many have coherence times of at least 2 h (not shown). Figure 4 is limited to 20 min because that is the limit imposed by experiment. It is important to analyze model and data the same way.

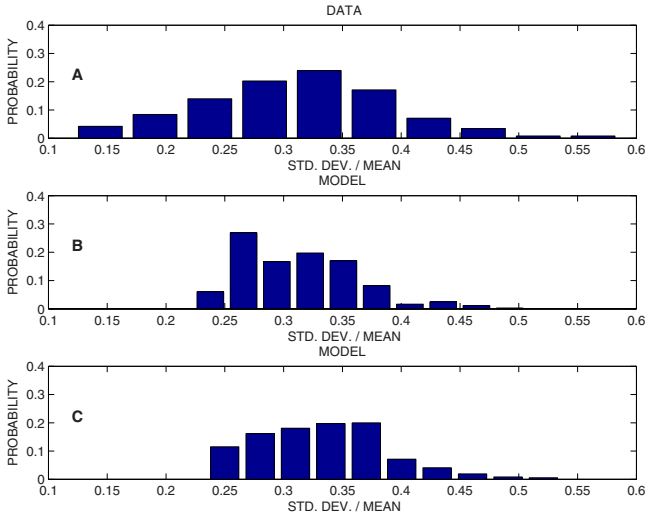


FIG. 7. (Color online) Same as Fig. 4 except for amplitude fluctuations. Within each window, mean and standard deviation of amplitude are computed. Their ratio yields f in Eq. (5), which gives values for the horizontal axis. Data are in A. Models for 20 and 25 dB signal-to-noise ratio are in panels B and C, respectively.

VI. AMPLITUDE FLUCTUATIONS

A. Data

Analysis of amplitude fluctuations is made in each of 380 windows discussed in Sec. V B. For each window, there are 42 realizations of peak amplitude, a_k, $k=1,2,3,\ldots,42$. We compute

$$f \equiv \mathrm{Std}[a_k]/\mathrm{Mean}[a_k], \tag{5}$$

where Std and Mean denote the standard deviation and the mean, respectively. This statistic is useful for computing channel capacity for some types of wireless acoustic modems in the sea.[5] We have 380 values of f, one for each window. An empirical probability distribution shows a peak near 0.35 with minimum and maximum values of 0.15 and 0.55 [Fig. 7(a)]. For informational purposes, we state that Eq. (5) is different than a scintillation index.

B. Model

Analysis of amplitude fluctuations is made in each of 365 windows discussed in Sec. V C. Noise is added to each of 150 records as before. For each window, we compute f from Eq. (5) except k goes to 150 model records. Empirical probability distribution of f differs somewhat from that derived from measurements [Fig. 7(b)]. Resemblance is better if we set record averaged peak signal-to-noise ratio to 25 dB [Fig. 7(c)]. Perhaps 25 dB better mimics average signal-to-noise ratio.

VII. DISCUSSION

Despite the fact that we have only one 20 min signal to work with, a means is found to increase reliability of statistical estimates by subdividing arriving energy into small windows with a duration of 0.03 s. Coherence times from individual windows are most likely near 20 min regardless of whether early or late energy is considered, or whether the window duration is 0.03 or 0.09 s. The model has many more degrees of freedom than the data. The model has snapshots of impulse response for 10 h at 4 min intervals.

Two issues seem to pertain to acoustic communication systems. First, they are often implemented by using differential Doppler to correct signals from different paths. In this experiment, it appears unnecessary to differentially correct paths for Doppler. A single Doppler correction at 27.28 s intervals is sufficient to yield coherence times between 2 and 19 min (Fig. 3). Second, the predicted values of probability distributions of coherence time and fractional amplitude are needed to predict channel capacity for some types of underwater modems.[5] Our predictions appear to be realistic enough to be useful for predicting channel capacity. Predicting the channel capacity is beyond the scope of this paper.

We do not know how to be sure what limits coherence time for signals lasting tens of minutes when receiver motion compels us to remove Doppler effects for each separate period of an M-sequence. In a separate experiment with the same source signal and different towed array in the same region, the observed and predicted coherence times agreed by correcting instrument motion for a single Doppler speed.[5] That array was towed at steadier speed. We will never know, but it does seem plausible that variable speed corrections used here do not affect the estimates of coherence time limited by the ocean. Effects of tides, other oscillations of the Pacific at shorter periods, and internal gravity waves are a factor of 10 too small to cause observed Doppler shifts (Secs. V A and V C). Receiver acceleration is the only phenomenon that could plausibly cause observed variation of Doppler. Whatever the cause, the corrected time series has a probability distribution of coherence time that looks very similar to that obtained from a standard model based on internal gravity waves alone.

It is reassuring to have a systematic method that yields coherence times similar to observed, at least here and in two other experiments checked so far.[4,5] Theories to date do not appear to be able to accurately predict coherence times.[4] If it turns out that Monte Carlo simulations of coherence time based on a standard spectrum of internal waves yield reliable coherence times in many experiments, a foundation will exist for further developing theory.

It seems possible that probability distribution for coherence time could extend past 20 min. We have not addressed whether coherence time exceeds 20 min since the hypothesis is untestable with a 20 min signal. What seems to be important is that modeled probability distribution for coherence time looks like that derived from data.

Finally, the Monte Carlo impulse responses are run on a supercomputer. Others are working on faster methods for implementing Monte Carlo approaches.[21]

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research Contract No. N00014-06-C-0031 and by a grant of computer time from the DOD High Performance Computing Modernization Program at the Naval Oceanographic Office. The ZRAY raytrace program was modified from the program

written by James Bowlin. I thank reviewers for their suggestions.

[1]B. M. Howe, S. G. Anderson, A. B. Baggeroer, J. A. Colosi, K. R. Hardy, D. Horwitt, F. W. Karig, S. Leach, J. A. Mercer, K. Metzger, L. O. Olson, D. A. Peckham, D. A. Reddaway, R. R. Ryan, R. P. Stein, K. von der Heydt, J. D. Watson, S. L. Weslander, and P. Worcester, "Instrumentation for the acoustic thermometry of ocean climate (ATOC) prototype Pacific Ocean network," Proceedings of the OCEANS 95 Conference Proceedings, San Diego, CA, 9–12 October 1995, pp. 1483–1500.

[2]C. Garrett and W. Munk, "Space-time scales of internal waves," Geophys. Fluid Dyn. **3**, 225–264 (1972).

[3]C. Garrett and W. Munk, "Space-time scales of internal waves: A progress report," J. Geophys. Res. **80**, 291–297 (1975).

[4]J. L. Spiesberger, F. Tappert, and A. R. Jacobson, "Blind prediction of broadband coherence time at basin-scales," J. Acoust. Soc. Am. **114**, 3147–3154 (2003).

[5]J. L. Spiesberger and D. Green, "Statistical characterization of very low frequency communication channels at ocean basin-scales," Proceedings of the MTS-IEEE Oceans '06 Conference, Boston, 2006, Paper No. 060322-06, pp. 1–6.

[6]J. L. Spiesberger, "Numerical prediction of coherent integration time at 75 Hz, 0.03 temporal resolution at 3250 km," Proceedings of the MTS-IEEE Oceans '06 Conference, Boston, 2006, Paper No. 060323-02, pp. 1–4.

[7]K. Wage, M. Dzieciuch, P. Worcester, B. Howe, and J. Mercer, "Mode coherence at megameter ranges in the North Pacific Ocean," J. Acoust. Soc. Am. **117**, 1565–1581 (2005).

[8]V. A. Del Grosso, "New equation for the speed of sound in natural waters with comparisons to other equations," J. Acoust. Soc. Am. **56**, 1084–1091 (1974).

[9]S. Levitus, "Climatological atlas of the world ocean," NOAA Professional Paper No. 13, U. S. Government Printing Office, Washington, DC, 1982.

[10]A. Ben-Menahem and S. J. Singh, *Seismic Waves and Sources* (Springer-Verlag, New York, 1981), p. 1108.

[11]J. A. Colosi and M. G. Brown, "Efficient numerical simulation of stochastic internal-wave-induced sound-speed perturbation fields," J. Acoust. Soc. Am. **103**, 2232–2235 (1998).

[12]M. A. Wolfson and J. L. Spiesberger, "Full wave simulation of the forward scattering of sound in a structured ocean: A comparison with observations," J. Acoust. Soc. Am. **106**, 1293–1306 (1999).

[13]V. Renard and J. P. Allenon, "Sea Beam, multi-beam echo-sounding in "Jean Charcot:" Description, evaluation and first results," Int. Hydrogr. Rev **56**, 35–67 (1979).

[14]Naval Oceanographic Office Data Model, Oceanographic and Atmospheric Master Library (OAML), Digital bathymetric database variable resolution (DBDB-V), Version 5.1, April, 2006.

[15]J. B. Bowlin, J. L. Spiesberger, and L. F. Freitag, "Ocean acoustical ray-tracing software RAY," Woods Hole Oceanographic Technical Report No. WHOI-93-10, 1992.

[16]J. L. Spiesberger, "An updated perspective on basin-scale tomography," J. Acoust. Soc. Am. **109**, 1740–1742 (2001).

[17]F. Tappert, J. L. Spiesberger, and L. Boden, "New full-wave approximation for ocean acoustic travel time predictions," J. Acoust. Soc. Am. **97**, 2771–2782 (1995).

[18]S. W. Golomb, *Shift Register Sequences* (Holden-Day, San Francisco, CA, 1967).

[19]J. L. Spiesberger, R. C. Spindel, and K. Metzger, "Stability and identification of ocean acoustic multipaths," J. Acoust. Soc. Am. **67**, 2011–2017 (1980).

[20]J. L. Spiesberger, P. J. Bushong, K. Metzger, and T. G. Birdsall, "Ocean acoustic tomography: Estimating the acoustic travel time with phase," IEEE J. Ocean. Eng. **14**, 108–119 (1989).

[21]A. K. Morozov and J. A. Colosi, "Stochastic differential equation analysis for ocean acoustic energy scattering by internal waves," J. Acoust. Soc. Am. **119**, 3344(2006).

Temporal coherence after multiple forward scattering through random three-dimensional inhomogeneities in an ocean waveguide

Tianrun Chen
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Purnima Ratilal
Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115

Nicholas C. Makris
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 23 January 2008; revised 17 June 2008; accepted 22 June 2008)

An analytical expression is derived for the temporal coherence of an acoustic field after multiple forward scattering through random three-dimensional (3D) inhomogeneities in an ocean waveguide. This expression makes it possible to predict the coherence time scale of field fluctuations in ocean-acoustic measurements from knowledge of the oceanography. It is used to explain the time scale of acoustic field fluctuations observed at megameter ranges in various deep ocean-acoustic transmission experiments. It is shown that this time scale is nonlinearly related to the much longer coherence time scale of deep ocean internal waves through a multiple forward scattering process. It is also shown that 3D scattering effects become pronounced when the acoustic Fresnel width exceeds the cross-range coherence length of the deep ocean internal waves, which lead to frequency and range-dependent power losses in the forward field that may help to explain historic long range measurements. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2967475]

PACS number(s): 43.30.Re [RCG] Pages: 2812–2822

I. INTRODUCTION

Internal waves cause random compressibility and density fluctuations that can have a pronounced accumulated effect on acoustic signals propagating over long ranges in the deep ocean. The resulting random multiple forward scattering causes significant fluctuations in the acoustic field,[1–8] leads to degradation in the temporal coherence of acoustic signals[9–11] and significant signal-dependent noise.[12] Understanding the properties of this signal-dependent noise is often critical to effectively employ acoustics to ocean remote sensing and communication,[13–19] as well as in the Acoustic Thermometry of Ocean Climate (ATOC).[10] Knowledge of the coherence time scale of a received signal is essential in (1) reducing the error of any measurement or estimate obtained from fluctuating acoustic field data by stationary averaging and (2) applying the fundamental coherent processing techniques of ocean acoustics, such as matched filtering, beamforming, matched-field and synthetic aperture processing. The coherence time scale, for example, is needed to determine the number of *statistically independent* samples of the received acoustic signal during a given measurement time, which can then be averaged to reduce estimation error and signal-dependent noise.[12,20] Since coherent processing must typically be restricted to *within* the coherence time scale of field fluctuations, a good estimate of the coherence time scale is usually necessary to design an effective experiment.

In this paper, we derive a general analytical expression for the temporal correlation function, which can be used to predict the coherence time scale of the acoustic power and field propagated through two-dimensional (2D) and three-dimensional (3D) random inhomogeneities, such as internal waves, bubbles, fish schools, eddies, and surface waves. From this expression, the coherence time scale is calculated for acoustic waves propagating through random internal wave inhomogeneities. It is shown that the coherence time scale of the acoustic field fluctuations observed at megameter ranges in various deep ocean-acoustic transmission experiments[10,21] can be explained by multiple forward scattering through linear internal waves in the deep ocean. The roughly 10 min acoustic coherence time scale measured[10,21] is shown to be nonlinearly related to the much longer 4 h coherence time scale of the internal waves. The latter is derived from the Garret–Munk (GM) spectrum.[22,23] Analysis of the temporal coherence function at zero time lag, which corresponds to the expected intensity of the fluctuating signal, shows that 3D scattering from internal waves can lead to power loss in the low frequency long range propagation in the deep ocean. For a given receiver range, these losses begin to become pronounced as the frequency decreases to the point where the Fresnel length[24] of the acoustic measurements exceeds the cross-range coherence length[24] of the internal waves. This leads to out-of-plane scattering that removes energy from the forward direction. After reaching a

 © 2008 Acoustical Society of America

maximum, power losses in the forward direction begin to decrease due to weakening Rayleigh–Born scattering as the acoustic frequency decreases. This may explain the unexpected attenuation observed in some historic long range (megameter) low frequency (5–75 Hz) measurements.[1,25]

Much related work in the 1970s and 1980s focused on using ray theory to work toward an estimate of temporal coherence by accumulating random phase fluctuations along isolated water-borne ray paths in the deep ocean.[2,26] Due to perceived oversimplifications, many moved away from ray theory and began performing numerical Monte Carlo simulations with the 2D parabolic wave equation. For example, in Ref. 27, a 2D Monte Carlo approach based on a parabolic approximation for a specific deep ocean environment was used to estimate the temporal coherence function. Some disadvantages of the Monte Carlo approach, however, are that (1) it does not provide a general analytical expression for the temporal coherence function but, instead, requires intensive numerical calculations for each specific case and (2) it is restricted to 2D propagation and scattering scenarios.

In Sec. II, analytical expressions for the temporal correlation functions of the acoustic power and field forward propagated through a slowly time-varying random medium are derived. In Sec. III, the mean and temporal coherence of the scatter function are derived for internal wave inhomogeneities in the deep ocean. This is required to determine the temporal coherence function of the acoustic power and field. Differences between 2D and 3D scattering processes are reviewed in Sec. IV. Illustrative examples are provided in Sec. V.

Before closing the Introduction, it is useful to first briefly review key steps in the derivation by Ratilal and Makris[24] that will also be needed here. Ratilal and Makris[24] combined the waveguide scattering theory and a differential marching procedure analogous to that used in free space optics by Rayleigh to derive the mean and power of the acoustic field forward propagated through 3D random inhomogeneities in terms of waveguide modes. That analysis is for the mean, spatial covariance, and temporal *variance* of the field. Since the temporal coherence function and the coherence time scale require the temporal *covariance*, it must be derived here, since the temporal variance and covariance are only equal at zero time lag. To obtain the temporal covariance, we follow the same marching procedure used in Ref. 24 to calculate the temporal coherence function of the acoustic power and forward field propagated through 3D inhomogeneities. We define the coherence time scale to be the e-folding time at which the temporal coherence function falls to $1/e$ of its zero time lag value. Knowledge of this time scale is often essential for designing underwater experiments.

Since the inhomogeneities are assumed to follow a stationary random process over time, the mean field remains as the time invariant result derived in Ref. 24. The process of propagating through a single differential slab of random inhomogeneities causes a change in the mean acoustic field. For each mode, this change can be expressed as a product of factors including the incident field, a complex modal wavenumber change induced by scattering, and the thickness of the slab. The mean field after multiple forward scattering through the inhomogeneous medium, obtained by integrating over differential slabs, takes the form of a product of the incident field and an exponential factor that involves the accumulated modal wavenumber changes over range from source to receiver. These wavenumber changes determine the dispersive and attenuating effects of the inhomogeneities, through their real and imaginary parts, respectively, and account for mode coupling in the scattering process.

Ratilal and Makris[24] used a similar marching procedure to derive the mean power of the forward field. The incremental change in acoustic power due to a single slab of inhomoegeneities can be expressed in terms of the depth integral of the second moment of the scattered field, as well as cross terms between the scattered and incident fields. This change can then be expressed as the product of the incident power, the difference between modal variance and attenuation coefficients, and slab thickness. Modal variance coefficients, introduced by depth integration of the scattered field second moment, depend on the variance of intrinsic properties of the inhomogeneities, such as compressibility and density. The acoustic power at the receiver range is then obtained by marching the power change equation from the source to the receiver through direct integration. This power is expressed as a product of the incident power and an exponential factor involving the range integration of the difference between the modal variance and attenuation coefficients.

II. ANALYTICAL MODEL FOR TEMPORAL COHERENCE OF THE ACOUSTIC POWER AND FIELD FORWARD PROPAGATED THROUGH A SLOWLY TIME-VARYING RANDOM MEDIUM

In this section, we derive an analytical expression for the temporal correlation function of the acoustic power and field forward propagated through a slowly time-varying random medium based on the normal mode method. We assume that (1) the random medium follows a stationary process and (2) the medium in a single slab is static during the time period of acoustic wave propagation through it. This is a valid assumption since the correlation time scale of medium fluctuation is much longer than the time it takes the acoustic wave to propagate through a single slab. Note that the derivation of the temporal correlation function of acoustic power (depth-integrated temporal correlation of acoustic total field) does not require an assumption of acoustic modal independence.[24] The assumption of modal independence is only used here to estimate the temporal correlation of the forward field at a specified receiver location $\mathbf{r}$. It is approximately valid when the random component of the field becomes a circular complex Gaussian random variable,[12] since the independence of at least a few dominant modes is then necessary by the central limit theorem. Gaussian field fluctuations are typical in many ocean-acoustic measurements.[12]

All the derivations in this section are based on a single frequency transmission. In the Appendix, we extend our model of the temporal correlation function to a narrow-band signal, which can be approximately expressed in terms of the temporal correlation function of the forward field.

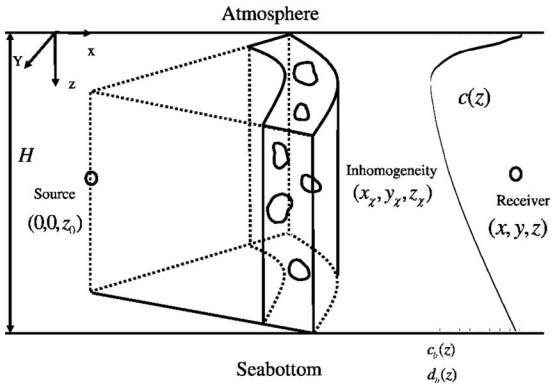


FIG. 1. The geometry of the source, receiver, and inhomogeneities.

A. Mean scattered field for a slowly time-varying random medium

Here, we provide the expression for the mean acoustic field scattered through slowly time-varying random inhomogeneities that are confined within a slab centered at ρ_s of thickness $\Delta\rho_s$.

As shown in Fig. 1, the origin of the coordinate system is placed at the air-water interface with the positive z-axis pointing downward. The source is located at the horizontal origin $\mathbf{r}_0=(0,0,z_0)$, receiver coordinates are given by $\mathbf{r}=(x,y,z)$, and inhomogeneity centers are given by $\mathbf{r}_\chi=(x_\chi,y_\chi,z_\chi)$. Here, t will solely be used to denote time dependence. [The notation t used in Ref. 24 is not for time dependence but denotes target coordinates $\mathbf{r_t}=(x_t,y_t,z_t)$.] Spatial cylindrical (ρ,ϕ,z) and spherical systems (r,θ,ϕ) are defined by $x=r\sin\theta\cos\phi$, $y=r\sin\theta\sin\phi$, $z=r\cos\theta$, and $\rho^2=x^2+y^2$. The horizontal and vertical wavenumber components for the nth mode are, respectively, $\xi_n=k\sin\alpha_n$ and $\gamma_n=k\cos\alpha_n$, where α_n is the elevation angle of the mode measured from the z-axis. Here, $0\le\alpha_n\le\pi/2$ so that the down- and upgoing plane wave components of each mode have elevation angles α_n and $\pi-\alpha_n$, respectively. The corresponding vertical wavenumber of the down- and upgoing components of the nth mode are γ_n and $-\gamma_n$, respectively, where $\Re\{\gamma_n\}\ge 0$. The azimuth angle of the mode is denoted by β. The wavenumber magnitude k equals the angular frequency ω divided by the sound speed c in the object layer, where $k^2=\xi_n^2+\gamma_n^2$. The geometry of spatial and wavenumber coordinates is shown in Fig. 2 of Ref. 28.

The scattered field from random inhomogeneities confined within the slab is found by summing contributions from unit volumes of inhomogeneity as described in Eq. (1) of Ref. 24. The scattered field discussed in the present paper is different from Eq. (1) of Ref. 24, since it varies as a function of time due to the slow *time-varying* random inhomogeneities in the slab according to

$$\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)=\int\int\int_{\Delta V_s}\varphi_s(\mathbf{r}|\mathbf{r}_0,\mathbf{r}_\chi,f,t)\mathrm{d}V_\chi, \quad (1)$$

where ΔV_s is the volume of the slab, $\varphi_s(\mathbf{r}|\mathbf{r}_0,\mathbf{r}_\chi,f,t)$ is the scattered field per unit volume of inhomogeneities *centered* at $\mathbf{r}_\chi$, and t is the time at the receiver $\mathbf{r}$.

The scattered field from a unit volume of slowly varying inhomogeneity, which has been derived from Green's theorem in Refs. 29 and 30, can be expressed as

$$\begin{aligned}\varphi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)=\sum_{m=1}\sum_{n=1}\frac{(4\pi)^2}{k}[&A_m(\mathbf{r}-\mathbf{r}_\chi)A_n(\mathbf{r}_\chi-\mathbf{r}_0)s_{\mathbf{r}_\chi,t}(\pi-\alpha_m,\beta_s(\phi,\phi_\chi);\alpha_n,\phi_\chi)\\&-B_m(\mathbf{r}-\mathbf{r}_\chi)A_n(\mathbf{r}_\chi-\mathbf{r}_0)s_{\mathbf{r}_\chi,t}(\alpha_m,\beta_s(\phi,\phi_\chi);\alpha_n,\phi_\chi)-A_m(\mathbf{r}-\mathbf{r}_\chi)B_n(\mathbf{r}_\chi-\mathbf{r}_0)\\&\times s_{\mathbf{r}_\chi,t}(\pi-\alpha_m,\beta_s(\phi,\phi_\chi);\pi-\alpha_n,\phi_\chi)+B_m(\mathbf{r}-\mathbf{r}_\chi)B_n(\mathbf{r}_\chi-\mathbf{r}_0)s_{\mathbf{r}_\chi,t}(\alpha_m,\beta_s(\phi,\phi_\chi);\pi-\alpha_n,\phi_\chi)],\end{aligned} \quad (2)$$

where $A_n(\mathbf{r}_\chi-\mathbf{r}_0)$ and $B_n(\mathbf{r}_\chi-\mathbf{r}_0)$ are the amplitudes of the *down-* and *upgoing* modal plane wave components incident on the inhomogeneity at $\mathbf{r}_\chi$ defined in Eqs. (3) and (4) of Ref. 24, $A_m(\mathbf{r}-\mathbf{r}_\chi)$ and $B_m(\mathbf{r}-\mathbf{r}_\chi)$ are the amplitudes of *up-* and *downgoing* modal components scattered from the inhomogeneity defined in Eqs. (5) and (6) of Ref. 24, $\beta_s(\phi,\phi_\chi)=\phi-\arcsin\{\rho_\chi/|\boldsymbol{\rho}-\boldsymbol{\rho}_\chi|\sin(\phi_\chi-\phi)\}$ is the receiver azimuth from the target, and $s_{\mathbf{r}_\chi,t}(\alpha,\beta;\alpha_i,\beta_i)$ is the scatter function density[24] of inhomogeneities at $\mathbf{r}_\chi$.

Assuming that the inhomogeneities obey a stationary process in time, the *mean* scatter function density is time invariant. The mean scattered field from a single inhomogeneous slab can then still be expressed as

$$\begin{aligned}\langle\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)\rangle&=\langle\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f)\rangle\\&=\sum_n\Phi_i^n(\mathbf{r}|\mathbf{r}_0,f)i\nu_n(\rho_s)\Delta\rho_s,\end{aligned} \quad (3)$$

by following Eq. (59) of Ref. 24, where $\nu_n(\rho_s)$ is the time-invariant change of the complex wavenumber due to scattering from the slab.

B. Marching temporal correlation through a slowly varying random waveguide with difference and integral equations

In this section, a difference equation is derived for the depth-integrated temporal correlation of the total field scattered from inhomogeneities confined within a slab centered at ρ_s. The depth-integrated temporal correlation of the total

field is then marched through the random waveguide to include multiple forward scattering from all inhomogeneities between the source and receiver.

The total field scattered from inhomogeneities within the slab is

$$\Phi(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)=\Phi_i(\mathbf{r}|\mathbf{r}_0,f)+\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t), \tag{4}$$

where

$$\Phi_i(\mathbf{r}|\mathbf{r}_0,f)=4\pi\frac{i}{d(z_0)\sqrt{8\pi}}e^{-i\pi/4}\sum_n u_n(z)u_n(z_0)\frac{e^{i\xi_n\rho}}{\sqrt{\xi_n\rho}} \tag{5}$$

is the incident field expressed in terms of acoustic normal modes.

The mean total field does not depend on t from Eq. (3) and, as in Eq. (83) of Ref. 24, is given by:

$$\begin{aligned}&\langle\Phi_T(\mathbf{r}|\mathbf{r}_0,f,t)\rangle\\&=\langle\Phi_T(\mathbf{r}|\mathbf{r}_0,f)\rangle\\&=4\pi\frac{i}{d(z_0)\sqrt{8\pi}}e^{-i\pi/4}\sum_n u_n(z)u_n(z_0)\frac{e^{i\xi_n\rho}}{\sqrt{\xi_n\rho}}e^{i\int_0^\rho\nu_n(\rho_s)\mathrm{d}\rho_s},\end{aligned} \tag{6}$$

where the addition of an exponential term accounts for the dispersion and attenuation induced by multiple scattering. From Eq. (4), the depth-integrated temporal correlation of the total field can be written as

$$\begin{aligned}&\int_0^\infty\frac{1}{d(z)}\langle\Phi(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)\Phi^*(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t')\rangle\mathrm{d}z\\&=\int_0^\infty\frac{1}{d(z)}|\Phi_i(\mathbf{r}|\mathbf{r}_0,f)|^2\mathrm{d}z\\&+\int_0^\infty\frac{1}{d(z)}(\langle\Phi_i(\mathbf{r}|\mathbf{r}_0,f)\Phi_s^*(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t')\rangle\\&+\langle\Phi_i^*(\mathbf{r}|\mathbf{r}_0,f)\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)\rangle)\mathrm{d}z\\&+\int_0^\infty\frac{1}{d(z)}\langle\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)\\&\times\Phi_s^*(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t')\rangle\mathrm{d}z.\end{aligned} \tag{7}$$

The first term at the right hand side of Eq. (7) is the incident intensity

$$\int_0^\infty\frac{1}{d(z)}|\Phi_i(\mathbf{r}|\mathbf{r}_0,f)|^2\mathrm{d}z=W_i(\boldsymbol{\rho}|\mathbf{r}_0)=\sum_n W_i^{(n)}(\boldsymbol{\rho}|\mathbf{r}_0), \tag{8}$$

where

$$W_i^{(n)}(\boldsymbol{\rho}|\mathbf{r}_0)=\frac{2\pi}{d^2(z_0)}|u_n(z_0)|^2\frac{1}{\boldsymbol{\rho}|\xi_n|}e^{-2\Im\{\xi_n\}\boldsymbol{\rho}}.$$

The second term of Eq. (7) is the depth-integrated cross term arising from the interference between the incident and the scattered field. By inserting the mean scattered field of Eq. (3) into the second term of Eq. (7) and by invoking modal orthogonality

$$\int_0^\infty\frac{1}{d(z)}u_m(z)u_n(z)\mathrm{d}z=\delta_{nm}, \tag{9}$$

integration leads to

$$\begin{aligned}&\int_0^\infty\frac{1}{d(z)}[\langle\Phi_i(\mathbf{r}|\mathbf{r}_0)\Phi_s^*(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t')\rangle\\&+\langle\Phi_i^*(\mathbf{r}|\mathbf{r}_0)\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)\rangle]\mathrm{d}z\\&=\int_0^\infty\frac{1}{d(z)}[\Phi_i(\mathbf{r}|\mathbf{r}_0)\langle\Phi_s^*(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t')\rangle+\Phi_i^*(\mathbf{r}|\mathbf{r}_0)\\&\times\langle\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)\rangle]\mathrm{d}z\\&=-\sum_n W_i^{(n)}(\boldsymbol{\rho}|\mathbf{r}_0)2\Im\{\nu_n(\rho_s)\}\Delta\rho_s.\end{aligned} \tag{10}$$

The last term of Eq. (7) is the depth-integrated temporal correlation of the field scattered from the slab. The temporal correlation of the scattered field $\langle\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)\times\Phi_s^*(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t')\rangle=\mathrm{corr}_{\Phi_s\Phi_s}(\Delta\rho_s(\rho_s),f,\tau=t-t')$ depends on the temporal and spatial correlation of the scatter function density $\langle s_{\mathbf{r}_\chi,t}(\alpha,\beta,\alpha_i,\beta_i)s^*_{\mathbf{r}_{\chi'},t'}(\alpha',\beta',\alpha'_i,\beta'_i)\rangle$, which will be discussed in Sec. II C. When $t=t'$, $\mathrm{corr}_{\Phi_s\Phi_s}(\Delta\rho_s(\rho_s),f,\tau)$ becomes the second moment of the scattered field $\langle|\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f)|^2\rangle$ of Eq. (61), Ref. 24. From Sec. III B of Ref. 24, the depth-integrated second moment of the scattered field is

$$\begin{aligned}&\int_0^\infty\frac{1}{d(z)}\langle|\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f)|^2\rangle\mathrm{d}z\\&=\sum_{n=1}^\infty W_i^{(n)}(\boldsymbol{\rho}|\mathbf{r}_0)\mu_n(\rho_s,\tau)|_{\tau=0}\Delta\rho_s,\end{aligned} \tag{11}$$

where $\mu_n(\rho_s,0)$ is the variance coefficient that contains a modal sum to account for modal coupling due to the random scattering process. When $t\neq t'$, we can apply the procedure used in the derivation of Eq. (11) to calculate the last term in Eq. (7),

$$\begin{aligned}&\int_0^\infty\frac{1}{d(z)}\langle\Phi_s(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)\Phi_s^*(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t')\rangle\mathrm{d}z\\&=\sum_{n=1}^\infty W_i^{(n)}(\boldsymbol{\rho}|\mathbf{r}_0)\mu_n(\rho_s,\tau=t-t')\Delta\rho_s,\end{aligned} \tag{12}$$

where $\mu_n(\rho_s,\tau)$ is the *temporal* variance coefficient, which accounts for modal coupling and quantifies the modal energy transferred from the mean field to the covariance field. It also describes how the forward field decorrelates as the time lag $\tau=t-t'$ increases after propagating through a slowly time-varying medium. The dependence of $\mu_n(\rho_s,\tau)$ on time lag τ is a consequence of its dependence on the temporal covariance of the scatter function density $\mathrm{cov}_{ss}(\rho_s,z_\chi,z_{\chi'},\tau)$, which will be discussed in Sec. II C.

Inserting Eqs. (8), (10), and (12) into Eq. (7), the depth-integrated temporal correlation of the total field is found to be

$$\int_0^\infty \frac{1}{d(z)}\langle \Phi(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t)\Phi^*(\mathbf{r}|\mathbf{r}_0,\Delta\rho_s(\rho_s),f,t')\rangle \mathrm{d}z$$
$$= \langle W_T(\boldsymbol{\rho}|\mathbf{r}_0,\tau)\rangle$$
$$= \sum_n W_i^{(n)}(\boldsymbol{\rho}|\mathbf{r}_0)(1+\mu_n(\rho_s,\tau)-2\Im(\nu_n(\rho_s)))\Delta\rho_s. \quad (13)$$

This can be rewritten as a difference equation

$$\Delta\langle W_T(\boldsymbol{\rho}|\mathbf{r}_0,\tau=t-t')\rangle = \sum_n \Delta\langle W_T^{(n)}(\boldsymbol{\rho}|\mathbf{r}_0,\tau)\rangle$$
$$= \sum_n W_i^{(n)}(\boldsymbol{\rho}|\mathbf{r}_0)(\mu_n(\rho_s,\tau)-2\Im(\nu_n(\rho_s)))\Delta\rho_s. \quad (14)$$

Following the marching process described in Sec. III B of Ref. 24, we have

$$\langle W_T(\boldsymbol{\rho}|\mathbf{r}_0,\tau)\rangle = \sum_n \langle W_T^{(n)}(\boldsymbol{\rho}|\mathbf{r}_0,\tau)\rangle$$
$$= \sum_n W_i^{(n)}(\boldsymbol{\rho}|\mathbf{r}_0)e^{\int_0^\rho(\mu_n(\rho_s,\tau)-2\Im\{\nu_n(\rho_s)\})\mathrm{d}\rho_s}. \quad (15)$$

Assuming independence between acoustic modes and following the derivation for the second moment of the forward field in Sec III C of Ref. 24, the temporal correlation of the forward field at receiver $\mathbf{r}$ can be expressed as

$$\langle \Phi(\mathbf{r}|\mathbf{r}_0,f,t)\Phi^*(\mathbf{r}|\mathbf{r}_0,f,t')\rangle$$
$$= \mathrm{corr}_{\Phi\Phi^*}(\mathbf{r}|\mathbf{r}_0,f,\tau)$$
$$= \sum_n |\phi_i^{(n)}(\mathbf{r}|\mathbf{r}_0)|^2 \exp\left(\int_0^\rho [\mu_n(\rho_s,\tau)-2\Im\{\nu_n(\rho_s)\}]\mathrm{d}\rho_s\right), \quad (16)$$

where $|\phi_i^{(n)}|^2$ is the incident intensity of acoustic mode n in a static ocean waveguide. Since both $\mu_n(\rho_s,\tau)$ and $\nu_n(\rho_s)$ depend on the position of inhomogeneities along the acoustic propagation path, range dependence in the scattering process can be taken into account.

C. Complex wavenumber change and temporal covariance coefficient

As explained in Sec. II A, the complex wavenumber change $\nu_n(\rho_s)$ is time independent. The expression for $\nu_n(\rho_s)$ is given in Eq. (60a) of Ref. 24 as

$$\nu_n(\rho_s) = \int_0^\infty \frac{2\pi}{k}\frac{1}{\xi_n}\frac{1}{d(z_t)}[(N_n^{(1)})^2 e^{i2\gamma_{nz_t}}$$
$$\times\langle s_{\mathbf{r}_{\chi',t}}(\pi-\alpha_n,\phi;\alpha_n,\phi)\rangle - N_n^{(2)}N_n^{(1)}$$
$$\times\langle s_{\mathbf{r}_{\chi',t}}(\alpha_n,\phi;\alpha_n,\phi)\rangle - N_n^{(1)}N_n^{(2)}$$
$$\times\langle s_{\mathbf{r}_{\chi',t}}(\pi-\alpha_n,\phi;\pi-\alpha_n,\phi)\rangle$$
$$+ (N_n^{(2)})^2 e^{-i2\gamma_{nz_t}}\langle s_{\mathbf{r}_{\chi',t}}(\alpha_n,\phi;\pi-\alpha_n,\phi)\rangle]\mathrm{d}z_t. \quad (17)$$

At zero time lag, the temporal covariance coefficient $\mu_n(\rho_s,\tau)$ becomes the variance coefficient $\mu_n(\rho_s,\tau=0)$. The variance coefficients for 2D and 3D scattering processes, respectively, are explicitly expressed in Eqs. (74) and (77) of Ref. 24. The variance coefficient contains a term $C_{s,s}(\rho_s,z_\chi,z_{\chi'},m,n)$ found in Eq. (72) of Ref. 24, which is a function of *spatial* correlation of the scatter function density $\mathrm{cov}_{ss}(s_{\rho_s,z_\chi},s_{\rho_s,z_{\chi'}})=\langle s_{\mathbf{r}_\chi}(\alpha,\beta;\alpha_i,\beta_i)s^*_{\mathbf{r}_{\chi'}}(\alpha',\beta';\alpha_i',\beta_i')\rangle$.

For nonzero time lag, the time dependence in $\mu_n(\rho_s,\tau)$ is introduced by the time lag dependent term $C_{s,s}(\rho_s,z_\chi,z_{\chi'},m,n,\tau)$, which is obtained by replacing $\mathrm{cov}_{ss}(s_{\rho_s,z_\chi},s_{\rho_s,z_{\chi'}})$ with $\mathrm{cov}_{ss}(\rho_s,z_\chi,z_{\chi'},\tau)$ in $C_{s,s}(\rho_s,z_\chi,z_{\chi'},m,n)$. The temporal covariance of the scatter function density $\mathrm{cov}_{ss}(\rho_s,z_\chi,z_{\chi'},\tau)$ is derived for 2D and 3D scenarios, respectively, in the following section. Then, we give the expressions of the $\mu_n(\rho_s,\tau)$ by following the derivation of $\mu(\rho_s,\tau=0)$ in Sec. III B of Ref. 24.

1. Inhomogeneities fully correlated within the Fresnel width (2D)

Here, the cross-range coherence length ℓ_y of the random inhomogeneity is greater than the Fresnel width[24] Y_F, which corresponds to an effective 2D scattering process for forward scatter. The spatio-temporal correlation of the scatter function density is given by

$$\langle s_{\mathbf{r}_{\chi',t}}(\alpha,\beta,\alpha_i,\beta_i)s^*_{\mathbf{r}_{\chi',t'}}(\alpha',\beta',\alpha_i',\beta_i')\rangle$$
$$\approx \ell_x(\rho_s,z_\chi,z_{\chi'})[\langle s_{\boldsymbol{\rho}_s,z_{\chi',t}}(\alpha,\beta,\alpha_i,\beta_i)s^*_{\boldsymbol{\rho}_s,z_{\chi',t'}}(\alpha',\beta',\alpha_i',\beta_i')\rangle$$
$$-\langle s_{\boldsymbol{\rho}_s,z_{\chi',t}}(\alpha,\beta,\alpha_i,\beta_i)\rangle\langle s^*_{\boldsymbol{\rho}_s,z_{\chi',t'}}(\alpha',\beta',\alpha_i',\beta_i')\rangle]\delta(x_\chi-x_{\chi'})$$
$$+\langle s_{\boldsymbol{\rho}_s,z_{\chi',t}}(\alpha,\beta,\alpha_i,\beta_i)\rangle\langle s^*_{\boldsymbol{\rho}_s',z_{\chi',t'}}(\alpha',\beta',\alpha_i',\beta_i')\rangle$$
$$= \ell_x(\rho_s,z_\chi,z_{\chi'})\mathrm{cov}_{ss}(\rho_s,z_\chi,z_{\chi'},\tau)\delta(x_\chi-x_{\chi'})$$
$$+\langle s_{\boldsymbol{\rho}_s,z_{\chi',t}}(\alpha,\beta,\alpha_i,\beta_i)\rangle\langle s^*_{\boldsymbol{\rho}_s',z_{\chi',t'}}(\alpha',\beta',\alpha_i',\beta_i')\rangle, \quad (18)$$

where ℓ_x is the coherence length[24] of the random inhomogeneity in the direction of the acoustic propagation. The temporal covariance of the scatter function density

$$\mathrm{cov}_{ss}(\rho_s,z_\chi,z_{\chi'},\tau=t-t')$$
$$= \langle s_{\boldsymbol{\rho}_s,z_{\chi',t}}(\alpha,\beta,\alpha_i,\beta_i)s^*_{\boldsymbol{\rho}_s,z_{\chi',t'}}(\alpha',\beta',\alpha_i',\beta_i')\rangle$$
$$-\langle s_{\boldsymbol{\rho}_s,z_{\chi',t}}(\alpha,\beta,\alpha_i,\beta_i)\rangle\langle s^*_{\boldsymbol{\rho}_s,z_{\chi',t'}}(\alpha',\beta',\alpha_i',\beta_i')\rangle \quad (19)$$

depends only on the time lag $\tau=t-t'$ not on the absolute time t and t' as a consequence of the temporal stationarity assumption for inhomogeneities in the slab.

The temporal variance coefficient only depends on ℓ_x under the 2D scattering scenario

$$\mu_n^{2\mathrm{D}}(\rho_s,\tau) = \sum_m \frac{1}{|\xi_m|}\int_0^\infty \mathrm{d}z_\chi \int_0^\infty \mathrm{d}z_{\chi'}\frac{\ell_x(\rho_s,z_\chi,z_{\chi'})}{\xi_m}$$
$$\times\frac{4\pi^2}{k(z_\chi)k(z_{\chi'})d(z_\chi)d(z_{\chi'})}C_{s,s}(\rho_s,z_\chi,z_{\chi'},m,n,\tau). \quad (20)$$

2. Inhomogeneities uncorrelated within the Fresnel width

Here, $\ell_y < Y_F$ so that inhomogeneities contained within the Fresnel width lead to a 3D scattering process and are uncorrelated. This leads to

$$\begin{aligned}
&\langle s_{\mathbf{r}_\chi,t}(\alpha,\beta,\alpha_i,\beta_i) s^*_{\mathbf{r}_{\chi'},t'}(\alpha',\beta',\alpha_i',\beta_i')\rangle \\
&\quad \approx A_c(\rho_s,z_\chi,z_{\chi'})[\langle s_{\boldsymbol{\rho}_s,z_\chi,t}(\alpha,\beta,\alpha_i,\beta_i) s^*_{\boldsymbol{\rho}_s,z_{\chi'},t'}(\alpha',\beta',\alpha_i',\beta_i')\rangle \\
&\quad - \langle s_{\boldsymbol{\rho}_s,z_\chi,t}(\alpha,\beta,\alpha_i,\beta_i)\rangle\langle s^*_{\boldsymbol{\rho}_s,z_{\chi'},t'}(\alpha',\beta',\alpha_i',\beta_i')\rangle]\delta(\boldsymbol{\rho}_\chi-\boldsymbol{\rho}_{\chi'}) \\
&\quad + \langle s_{\boldsymbol{\rho}_s,z_\chi,t}(\alpha,\beta,\alpha_i,\beta_i)\rangle\langle s^*_{\boldsymbol{\rho}'_s,z_{\chi'},t'}(\alpha',\beta',\alpha_i',\beta_i')\rangle \\
&\quad = A_c(\rho_s,z_\chi,z_{\chi'})\mathrm{cov}_{ss}(\rho_s,z_\chi,z_{\chi'},\tau)\delta(\boldsymbol{\rho}_\chi-\boldsymbol{\rho}_{\chi'}) \\
&\quad + \langle s_{\boldsymbol{\rho}_s,z_\chi,t}(\alpha,\beta,\alpha_i,\beta_i)\rangle\langle s^*_{\boldsymbol{\rho}'_s,z_{\chi'},t'}(\alpha',\beta',\alpha_i',\beta_i')\rangle, \qquad (21)
\end{aligned}$$

where $A_c(\rho_s,z_\chi,z_{\chi'})$ is the coherence area of the inhomogeneities.[24]

The temporal covariance is a function of $A_c(\rho_s,z_\chi,z_{\chi'})$ under the 3D scattering scenario

$$\begin{aligned}
\mu_n^{3\mathrm{D}}(\rho_s,\tau) = \sum_m \sqrt{\frac{\rho}{2\pi\xi_m\rho_s(\rho-\rho_s)}}\frac{1}{|\xi_m|}\int_0^\infty \mathrm{d}z_\chi \int_0^\infty \mathrm{d}z_{\chi'} \\
\times A_c(\rho_s,z_\chi,z_{\chi'})\frac{4\pi^2}{k(z_\chi)k(z_{\chi'})d(z_\chi)d(z_{\chi'})} \\
\times C_{s,s}(\rho_s,z_\chi,z_{\chi'},m,n,\tau). \qquad (22)
\end{aligned}$$

In summary, the spatio-temporal correlation of the scatter function density $\langle s_{\mathbf{r}_\chi,t}(\alpha,\beta;\alpha_i,\beta_i) \times s^*_{\mathbf{r}_{\chi'},t'}(\alpha',\beta';\alpha_i',\beta_i')\rangle$ is expressed as a function of the temporal covariance of the scatter function density $\mathrm{cov}_{ss}(\rho_s,z_\chi,z_{\chi'},\tau)$, which accounts for slow time variations of inhomogeneities via time lag τ.

III. STATISTICS OF THE SCATTER FUNCTION DENSITY OF INTERNAL WAVE INHOMOGENEITIES IN A DEEP OCEAN ENVIRONMENT

In this section, we derive expressions for the mean and temporal correlation of the scatter function density when the random inhomogeneities are internal waves in a deep ocean environment. The scatter function density for a coherent volume of internal wave inhomogeneity centered at the horizontal location $\boldsymbol{\rho}_s$ is expressed in terms of the compressibility fraction Γ_κ and the density fraction Γ_ρ by applying the Rayleigh–Born approximation to Green's theorem, as shown in Sec. B of Ref. 31,

$$\begin{aligned}
s_{\boldsymbol{\rho}_s,z_\chi,t}(\alpha,\beta,\alpha_i,\beta_i) = \frac{1}{A_c(\rho_s,z_\chi)}\int\int_{A_c}\frac{k^3}{4\pi}[\Gamma_\kappa(\boldsymbol{\rho}_\chi,z_\chi,t) \\
+ \eta(\mathbf{k},\mathbf{k}_i)\Gamma_d(\boldsymbol{\rho}_\chi,z_\chi,t)]e^{i(\xi_i-\xi)\cdot\mathbf{u}_\chi}\mathrm{d}^2\mathbf{u}_\chi, \qquad (23)
\end{aligned}$$

where $\boldsymbol{\rho}_\chi=\boldsymbol{\rho}_s+\mathbf{u}_\chi$ and $\eta(\mathbf{k},\mathbf{k}_i)=\mathbf{k}_i\cdot\mathbf{k}/k^2=\cos\alpha_i\cos\alpha+\sin\alpha_i\sin\alpha\cos(\beta_i-\beta)$ is the cosine of the angle between the incident and scattered plane wave directions.

The mean scatter function density is

$$\begin{aligned}
\langle s_{\boldsymbol{\rho}_s,z_\chi,t}(\alpha,\beta,\alpha_i,\beta_i)\rangle &= \frac{1}{A_c(\rho_s,z_\chi)}\int\int_{A_c}\frac{k^3}{4\pi}\langle\Gamma_\kappa(\boldsymbol{\rho}_\chi,z_\chi,t) \\
&\quad + \eta(\mathbf{k},\mathbf{k}_i)\Gamma_d(\boldsymbol{\rho}_\chi,z_\chi,t)\rangle e^{i(\xi_i-\xi)\cdot\mathbf{u}_\chi}\mathrm{d}^2\mathbf{u}_\chi \\
&= \frac{1}{A_c(\rho_s,z_\chi)}\frac{k^3}{4\pi}\langle\Gamma_\kappa(\boldsymbol{\rho}_\chi,z_\chi,t) \\
&\quad + \eta(\mathbf{k},\mathbf{k}_i)\Gamma_d(\boldsymbol{\rho}_\chi,z_\chi,t)\rangle\int\int_{A_c} e^{i(\xi_i-\xi)\cdot\mathbf{u}_\chi}\mathrm{d}^2\mathbf{u}_\chi \qquad (24)
\end{aligned}$$

and the temporal correlation of the scatter function is

$$\begin{aligned}
&\langle s_{\boldsymbol{\rho}_s,z_\chi,t}(\alpha,\beta,\alpha_i,\beta_i) s^*_{\boldsymbol{\rho}_s,z_{\chi'},t'}(\alpha',\beta',\alpha_i',\beta_i')\rangle \\
&\quad = \frac{1}{A_c(\rho_s,z_\chi)A_c(\rho_s,z_{\chi'})}\int\int_{A_c}\int\int_{A_c'} e^{i(\xi_i-\xi)\cdot(\mathbf{u}_\chi-\mathbf{u}_{\chi'})} \\
&\quad \times \mathrm{cov}_{\mathcal{FF}}(\boldsymbol{\rho}_\chi,\boldsymbol{\rho}_{\chi'},z_\chi,z_{\chi'},t,t')\mathrm{d}^2\mathbf{u}_\chi\mathrm{d}^2\mathbf{u}_{\chi'}, \qquad (25)
\end{aligned}$$

where

$$\begin{aligned}
&\mathrm{cov}_{\mathcal{FF}}(\boldsymbol{\rho}_\chi,\boldsymbol{\rho}_{\chi'},z_\chi,z_{\chi'},t,t') \\
&\quad = \left(\frac{k^3}{4\pi}\right)^2\langle[\Gamma_\kappa(\boldsymbol{\rho}_\chi,z_\chi,t)+\eta(\mathbf{k},\mathbf{k}_i)\Gamma_d(\boldsymbol{\rho}_\chi,z_\chi,t)] \\
&\quad \times[\Gamma_\kappa(\boldsymbol{\rho}_{\chi'},z_{\chi'},t')+\eta(\mathbf{k},\mathbf{k}_i)\Gamma_d(\boldsymbol{\rho}_{\chi'},z_{\chi'},t')]\rangle \qquad (26)
\end{aligned}$$

is the spatio-temporal correlation of intrinsic scattering properties. To calculate Eqs. (24) and (25), the statistical moments of fractional changes in compressibility Γ_κ and density Γ_d are required.

Since the fluctuations of the sound speed (Δc) and density (Δd) arising from random internal waves are much smaller than the unperturbed or local equilibrium sound speed and density, the fractional change of compressibility Γ_κ and density Γ_d can be expanded up to second order in Taylor series

$$\begin{aligned}
\Gamma_\kappa &\approx \left[-\frac{2\Delta c}{c_0}-\frac{\Delta d}{d_0}\right]+\left[3\left(\frac{\Delta c}{c_0}\right)^2+\left(\frac{\Delta d}{d_0}\right)^2+2\frac{\Delta c\Delta d}{c_0 d_0}\right], \\
\Gamma_d &\approx \frac{\Delta d}{d_0}-\left(\frac{\Delta d}{d_0}\right)^2. \qquad (27)
\end{aligned}$$

Fluctuations of the sound speed and density, for practical purposes, are linearly dependent on the displacement of the internal wave[2,32] $\xi(\boldsymbol{\rho}_\chi,z_\chi,t)$ via

$$\begin{aligned}
\frac{\Delta c(\boldsymbol{\rho}_\chi,z_\chi,t)}{c_0} &= \xi(\boldsymbol{\rho}_\chi,z_\chi,t)\bar{G}(z_\chi)n^2(z_\chi), \\
\frac{\Delta d(\boldsymbol{\rho}_\chi,z_\chi,t)}{d_0} &= \xi(\boldsymbol{\rho}_\chi,z_\chi,t)g^{-1}n^2(z_\chi), \qquad (28)
\end{aligned}$$

where $\bar{G}(z_\chi)$ is a function of the potential temperature and salinity,[2] g is the gravitational constant, and $n(z_\chi)$ is the buoyancy frequency or *Brunt–Väisälä* frequency

$$n^2(z_\chi) = -gd^{-1}\frac{\partial d_p(z_\chi)}{\partial z_\chi}, \tag{29}$$

where $d_p(z_\chi)$ is the potential density.[2]

The displacement ξ of internal waves at location $(\boldsymbol{\rho}_\chi, z_\chi)$ and time t is taken to be a zero-mean Gaussian random variable[2]

$$\xi(\boldsymbol{\rho}_\chi, z_\chi, t) = \sum_j \int H(\boldsymbol{\sigma}, j) W_j(\sigma, z_\chi) \exp(i[\boldsymbol{\sigma}\cdot\boldsymbol{\rho}_\chi - \Omega(\sigma, j)t])\mathrm{d}^2\boldsymbol{\sigma}, \tag{30}$$

where $H(\boldsymbol{\sigma}, j)$ is a zero-mean Gaussian random variable that specifies the jth modal amplitude of internal waves at wave number $\boldsymbol{\sigma}$ and $W_j(\sigma, z_\chi)$ is the jth modal shape of internal waves at depth z_χ. Internal wave angular frequency Ω is related to the magnitude of the internal wave wavenumber σ via the dispersion relation given in Ref. 2.

Assuming that the internal wave field follows a stationary random process in the horizontal space and in time, the spatio-temporal covariance of internal wave displacements at two measurement points $\mathbf{r}_\chi(\boldsymbol{\rho}_\chi, z_\chi)$ and $\mathbf{r}'_\chi(\boldsymbol{\rho}_{\chi'}, z_{\chi'})$ and two times t and t' can be expressed as a function of the horizontal separation $\mathbf{R}=\boldsymbol{\rho}_\chi - \boldsymbol{\rho}'_\chi$ and time lag $\tau = t - t'$;[2,22,23] we have

$$\begin{aligned}\mathrm{cov}_{\xi\xi}(\mathbf{R}, \tau, z_\chi, z_{\chi'}) &= \langle \xi(\boldsymbol{\rho}_\chi, z_\chi, t)\xi(\boldsymbol{\rho}_{\chi'}, z_{\chi'}, t')\rangle \\ &\quad - \langle \xi(\boldsymbol{\rho}_\chi, z_\chi, t)\rangle\langle \xi(\boldsymbol{\rho}_{\chi'}, z_{\chi'}, t')\rangle \\ &= \sum_j \int F_j(\boldsymbol{\sigma}) W_j(\sigma, z_\chi) W_j^*(\sigma, z_{\chi'}) \\ &\quad \times \exp(i[\boldsymbol{\sigma}\cdot\mathbf{R} - \omega(\sigma, j)\tau])\mathrm{d}^2\boldsymbol{\sigma}, \end{aligned} \tag{31}$$

where $F_j(\boldsymbol{\sigma}) = \langle |H(\boldsymbol{\sigma}, j)|^2\rangle$ is the GM spectrum of internal waves, whose expressions and parameters can be found in p. 56 of Ref. 2.

After inserting Eq. (28) into Eq. (27), and Eq. (27) into Eq. (26), the spatio-temporal covariance of intrinsic scattering properties becomes

$$\begin{aligned}&\mathrm{cov}_{\mathcal{F}\mathcal{F}}(\boldsymbol{\rho}_\chi, \boldsymbol{\rho}'_\chi, z_\chi, z_{\chi'}, t, t') \\ &= \frac{k^6}{(4\pi)^2} h(z_\chi) h(z_{\chi'}) \langle \xi(\boldsymbol{\rho}_\chi, z_\chi, t)\xi^*(\boldsymbol{\rho}'_\chi, z_{\chi'}, t')\rangle \\ &= \frac{k^6}{(4\pi)^2} h(z_\chi)(z_{\chi'}) \mathrm{cov}_{\xi\xi}(\mathbf{R}, \tau, z_\chi, z_{\chi'}), \end{aligned} \tag{32}$$

where $h(z_\chi) = [2\bar{G}(z_\chi) + (1-\eta)/g]n^2(z_\chi)$.

We approximate the displacements of internal waves at two horizontal positions $\boldsymbol{\rho}_\chi$ and $\boldsymbol{\rho}_{\chi'}$ within the coherence area A_c as being fully correlated[24,31] such that

$$\begin{aligned}&\mathrm{cov}_{\mathcal{F}\mathcal{F}}(\boldsymbol{\rho}_\chi, \boldsymbol{\rho}'_\chi, z_\chi, z_{\chi'}, t, t') \\ &\approx \frac{k^6}{(4\pi)^2} h(z_\chi) h(z_{\chi'}) \mathrm{cov}_{\xi\xi}(0, \tau, z_\chi, z_{\chi'}). \end{aligned} \tag{33}$$

The mean scatter function density is found to be proportional to the second moment of the internal wave displacement from Eqs. (24), (27), and (30), such that Eq. (19) becomes

$$\begin{aligned}&\mathrm{cov}_{ss}(\rho_s, z_\chi, z_{\chi'}, \tau) \\ &\approx \langle s_{\boldsymbol{\rho}_s, z_\chi, t}(\alpha, \beta, \alpha_i, \beta_i) s^*_{\boldsymbol{\rho}_s, z_{\chi'}, t'}(\alpha', \beta', \alpha'_i, \beta'_i)\rangle. \end{aligned} \tag{34}$$

This is because the *square* of the mean scatter function density proportional to the fourth power of internal wave displacement, is much smaller than $\langle s_{\boldsymbol{\rho}_s, z_\chi, t} s^*_{\boldsymbol{\rho}_s, z_{\chi'}, t'}\rangle$, which is on the order of the second moment of internal wave displacement.

Equation (34) can be expressed as

$$\begin{aligned}\mathrm{cov}_{ss}(\rho_s, z_\chi, z_{\chi'}, \tau) &\approx \frac{k^6 h(z_\chi) h(z_{\chi'})}{(4\pi)^2 A_c(z_\chi) A_c(z_{\chi'})} \mathrm{cov}_{\xi\xi}(0, \tau, z_\chi, z_{\chi'}) \\ &\quad \times \int\int_{A_c} \int\int_{A'_c} e^{i(\xi_i - \xi)\cdot(\mathbf{u}_\chi - \mathbf{u}_{\chi'})} \mathrm{d}^2\mathbf{u}_\chi \mathrm{d}^2\mathbf{u}'_\chi \end{aligned} \tag{35}$$

by substituting Eq. (31) into Eq. (33), then Eq. (33) into Eq. (25). By replacing $\mathrm{cov}_{ss}(s_{\rho_s, z_\chi}, s_{\rho_s, z_{\chi'}})$ with $\mathrm{cov}_{ss}(\rho_s, z_\chi, z_{\chi'}, \tau)$ in Eq. (72) of Ref. 24, we have

$$\begin{aligned}&C_{s,s}(\rho_s, z_\chi, z_{\chi'}, m, n, \tau) \\ &= u_n(z_\chi) u_n^*(z_{\chi'}) u_m(z_\chi) u_m^*(z_{\chi'}) \mathrm{cov}_{ss}(\rho_s, z_\chi, z'_\chi, \tau), \end{aligned} \tag{36}$$

which leads to $\mu_n(\rho_s, \tau)$ by inserting Eq. (36) into Eq. (20).

A purely real modal horizontal wavenumber change $\nu_n(\rho_s)$ is obtained by substituting the mean scatter function density of Eq. (24) into Eq. (60a) cf. Ref. 24. This only accounts for dispersion but not for attenuation in the mean forward field. Assuming no power loss within the Fresnel region or in the forward direction for a 2D scattering process, the depth-integrated intensity at zero time lag $\langle W_T(\boldsymbol{\rho}|\mathbf{r}_0, \tau = 0)\rangle$ must equal the depth-integrated incident intensity $\Sigma_n W_i^{(n)}$ of Eq. (15). This requires

$$\Im(\nu_n^{2\mathrm{D}}(\rho_s)) = \tfrac{1}{2}\mu_n^{2\mathrm{D}}(\rho_s, \tau = 0). \tag{37}$$

Out-of-plane scattering becomes important in 3D scenarios and leads to power loss in the forward direction, which requires an imaginary part in $\nu_n(\rho_s)$. We apply the waveguide extinction theorem[30,33] to calculate $\Im(\nu_n(\rho_s))$ in Eq. (36) of Ref. 28.

IV. 2D AND 3D SCATTERING PROCESSES

The Fresnel width in cross range[24] is defined to be where the incident and forward scattered fields are highly coherent and have a phase difference of less than $\pi/4$. The Fresnel width, $Y_F(\rho, \rho_s) = \sqrt{[\lambda(\rho - \rho_s)\rho_s]/\rho}$, depends on the range of the source, receiver, and scatterer,[24,31] where λ is the acoustic wavelength. The maximum Fresnel width[24,31] $Y_{F_{\max}}(\rho, \rho_s) = \sqrt{\lambda\rho/4}$ occurs at the midpoint between the source and the receiver. When $Y_F < \ell_y$, an internal wave inhomogeneity is correlated within the Fresnel width, which leads to an effective 2D scattering process.

As the receiver range increases, Y_F exceeds ℓ_y, 3D scattering initiates, and uncorrelated internal-wave inhomogene-

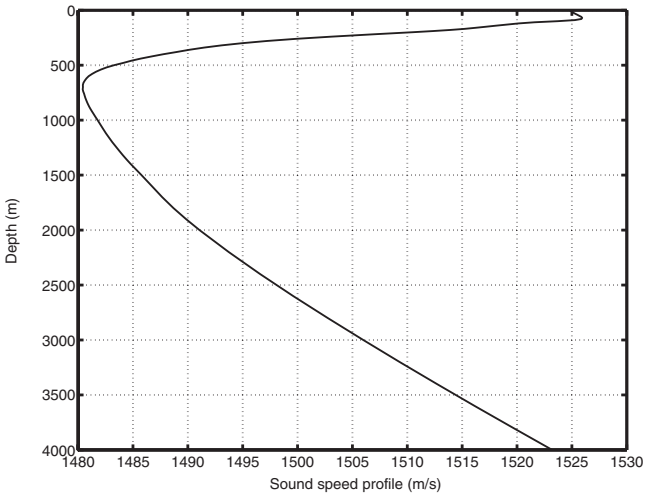


FIG. 2. Representative sound speed profile for the Pacific ocean region between latitudes 20°N and 31°N and longitudes 123°W and 154°W. The depth of the sound-channel axis is roughly 700 m.

ities appear within the Fresnel region. This leads to out-of-plane scattering that causes additional power loss in the forward direction.

V. ILLUSTRATIVE EXAMPLES

Here, a water column of $H=4000$ m depth is used to simulate the geometry of a deep ocean waveguide. The bottom sediment half-space is composed of sand with density $d_b=1.9$ g/cm^3 and sound speed $c_b=1700$ m/s. The attenuation coefficients in the water column and bottom are $\alpha=6\times10^{-5}$ dB/λ and $\alpha_b=0.8$ dB/λ, respectively. A point source transmits acoustic waves at a frequency of 75 Hz. Both the source and the point receiver are located at a depth of 1000 m. We use a sound speed profile calculated from the historical temperature and salinity data[10] corresponding to the Pacific ocean region between latitudes 20 °N and 31 °N and longitudes 123 °W and 154 °W, as shown in Fig. 2.

The temporal correlation function of internal wave displacements at the depth of the sound channel axis is shown in Fig. 3 following Eq. (31). The e-folding correlation time scale of the internal wave field is seen to be approximately 4 h at that depth. As the water depth increases, the coherence time scale of internal waves also increases because internal wave displacements decrease. The temporal correlation of the depth-integrated intensity and the acoustic forward field are shown in Fig. 4 following Eqs. (15) and (16), respectively. Uncertainty in the internal wave energy level leads to variations in the e-folding correlation time scale of the depth-integrated intensity and the forward scattered acoustic field. At three typical GM internal wave energy levels, for example, the e-folding correlation time scales vary between 7 and 14 min, as also shown in Fig. 4. These time scales are more than an order of magnitude smaller than that of the internal waves. Scattering from a single slab containing internal waves causes only a small change in the temporal correlation of the depth-integrated intensity. This is due to the weak scattering from a single slab and the very long time scale of the internal waves with respect to acoustic travel

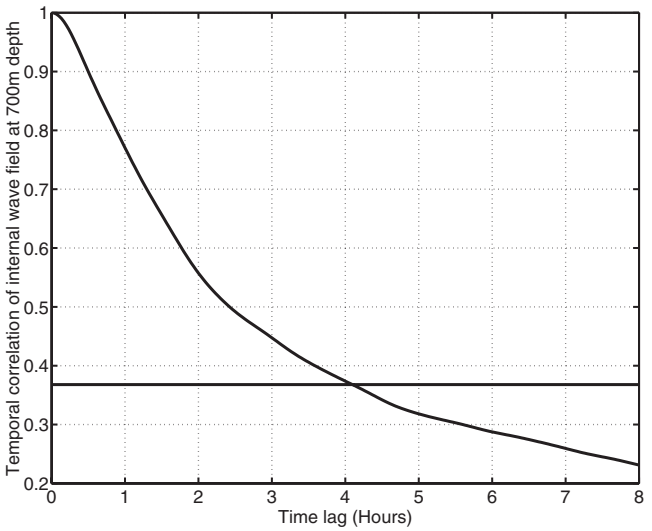


FIG. 3. Temporal correlation function of the internal-wave displacement at 700 m as a function of time lag. The temporal correlation is the *normalized* spatio–temporal covariance of Eq. (28) at zero horizontal separation. The e-folding coherence time is approximately 4 h. The solid horizontal line is plotted at the correlation function value $1/e$.

time through the slab. As acoustic power, or the depth-integrated intensity, is propagated through a series of uncorrelated slabs of inhomogeneities over range, *accumulated* multiple scattering dramatically degrades temporal correlation. This leads to coherence time scales for the acoustic power and forward field that are much shorter than that of the internal wave field. An assumption of modal independence is not required to obtain the temporal correlation and the time scale of the acoustic power.

Acoustic power loss due to scattering is

$$\mathrm{PL}(\boldsymbol{\rho}|\mathbf{r}_0) = 10\log\langle W_T(\boldsymbol{\rho}|\mathbf{r}_0,\tau=0)\rangle - 10\log W_i(\boldsymbol{\rho}|\mathbf{r}_0), \tag{38}$$

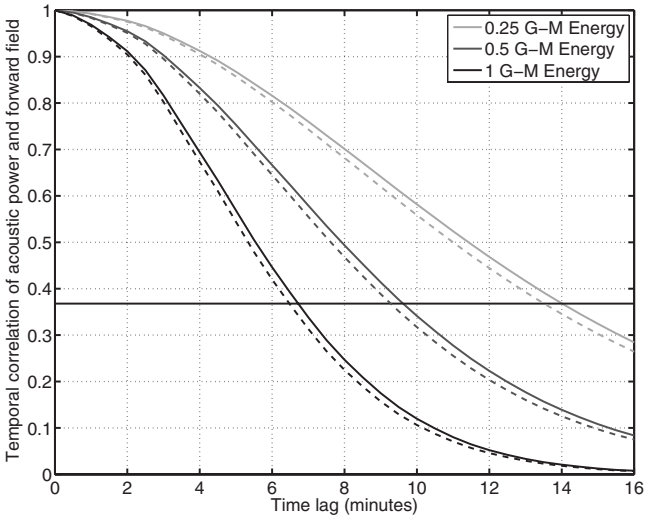


FIG. 4. Temporal correlation function of acoustic power and forward field for 3250 km source-receiver separation as a function of time lag for various possible GM energy levels. The temporal correlations of acoustic power and forward field are the *normalized* temporal covariances of Eqs. (17) and (18), respectively. The e-folding coherence time varies between 7 and 14 min depending on the internal wave energy level. The temporal correlation functions of the acoustic power and forward field are plotted with dashed and solid lines, respectively. The solid horizontal line is plotted at the correlation function value $1/e$.

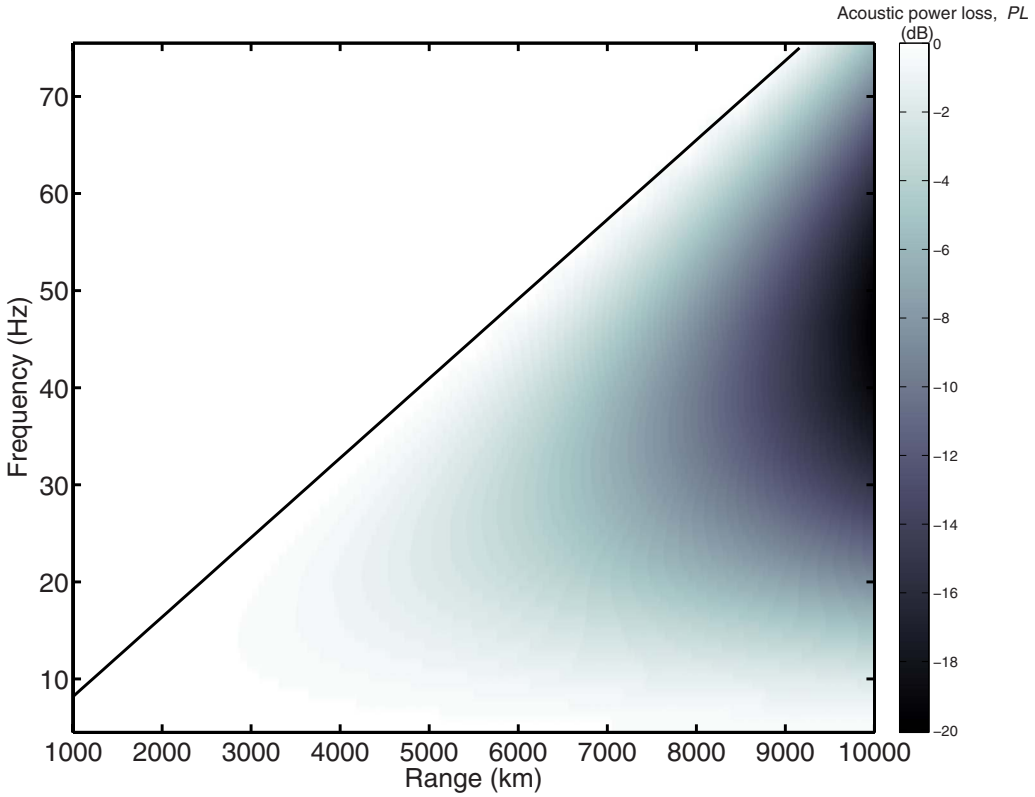


FIG. 5. (Color online) Total acoustic power loss in forward propagation through deep ocean random internal waves as a function of frequency and the source-receiver separation from Eq. (38). The 2D scattering region, $\ell_y > Y_F$, is above the black line and the 3D scattering region, $\ell_y < Y_F$, is below the black line.

where $\langle W_T(\boldsymbol{\rho}|\mathbf{r}_0, \tau=0)\rangle$ is the total acoustic power of Eq. (15) at zero time lag and $W_i(\boldsymbol{\rho}|\mathbf{r}_0)$ is the incident acoustic power of Eq. (8). We plot the acoustic power loss as a function of receiver range and acoustic frequency in Fig. 5. The 2D and 3D scattering regions are separated by a black line showing where $(Y_F)_{\max} = \ell_y$. In the 2D region, there is no power loss in the forward direction, as assumed in Sec. IV. In the 3D region, power loss monotonically increases with source-receiver separation for fixed frequencies. At a given source-receiver separation, however, power loss reaches a maximum in the low frequency regime. This is because two competing factors determine the total power loss: Y_F and $\mu_n^{3D}(\rho_s, 0)$. As frequency decreases, Y_F becomes larger leading to more uncorrelated internal wave inhomogeneities in the Fresnel region. This leads to more out-of-plane scattering and power loss. However, the f^{-6} Rayleigh–Born scattering roll-off for any individual inhomogeneity results in an $f^{-2.5}$ roll-off in variance coefficient $\mu_n(\rho_s, 0)$ at zero time lag, as seen in Eq. (15). This leads to less power loss.

VI. CONCLUSIONS

We have derived an analytical expression for the temporal covariance of the depth-integrated intensity or the acoustic power propagated through 3D random inhomogeneities, from which the coherence time scale of field or power fluctuations can be determined. Knowledge of the coherence time scale is typically essential in ocean-acoustic remote sensing. This is because it determines (1) the number of fluctuations in a given measurement period, which determines the amount of error reduction possible by stationary averaging in any ocean-acoustic remote sensing measurement, and (2) the time window within which the coherent processing essential to ocean-acoustic remote sensing, such as matched filtering, beamforming, matched field processing, and synthetic aperture processing, can be conducted. We have provided a general and rapid way of estimating the coherence time scale to aid in the design of ocean-acoustic experiments and the interpretation of experimental measurements. We show by analysis that the time scale of acoustic power fluctuations after megameter range propagation through internal waves in the deep ocean is roughly 10 min, which matches experimental measurements, and is more than an order of magnitude smaller than the 240 min correlation time scale of the internal wave field. This discrepancy between the acoustic and internal wave time scales is explained by the present theory as the accumulated effect of multiple forward scattering through internal waves on acoustic waves.

We find that power loss due to 3D scattering from internal wave inhomogeneities becomes pronounced when the Fresnel width exceeds the cross-range correlation length of internal waves. As source-receiver separation increases, the Fresnel width increases and contains more uncorrelated internal wave inhomogeneities, which leads to a monotonic increase in power loss. For a given source-receiver separation and decreasing frequency, power loss first increases as the internal wave incoherence accumulates within the Fresnel width, and then decreases due to weakening Rayleigh–Born scattering for any individual inhomogeneity. This may explain the unexpectedly high attenuations historically observed below 100 Hz.[1,25]

APPENDIX: TEMPORAL COHERENCE OF NARROW-BAND ACOUSTIC SIGNALS

Here we discuss extension of the single-frequency transmission of the main text to narrow-band signals. Let the acoustic signal measured at a receiver be

$$\Psi(r|r_0,t) = \Psi_i(r|r_0,t) + \Psi_s(r|r_0,t), \tag{A1}$$

where $\Psi_i(r|r_0,t) = \int_{-\infty}^{\infty} Q(f)\Phi_i(r|r_0,f)e^{-i2\pi ft}df$ is the incident field, $\Psi_s(r|r_0,t) = \int_{-\infty}^{\infty} Q(f)\Phi_s(r|r_0,f,t)e^{-i2\pi ft}df$ is the scattered field, and $Q(f)$ is the source spectrum. The temporal coherence or autocorrelation function of the acoustic signal is

$$\begin{aligned}\langle \Psi(r|r_0,t)\Psi^*(r|r_0,t')\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Q(f)Q(f') \\ &\times \langle \Phi(r|r_0,f,t)\Phi(r|r_0,f',t')\rangle \\ &\times e^{-i2\pi(ft-f't')}\mathrm{d}f\mathrm{d}f'. \end{aligned}\tag{A2}$$

In order to evaluate Eq. (A2), we need to calculate the temporal correlation of the total field $\langle \Phi(r|r_0,f,t)\Phi^*(r|r_0,f',t')\rangle$ at two different frequencies f and f'.

Derivations for the depth-integrated second moment of the scattered field in Sec. IV B of Ref. 24 and the second term of Eq. (7) in Sec. II B of the present paper rely on modal orthogonality

$$\int_0^{\infty} \frac{1}{d(z)} u_m(f,z)u_n(f,z)\mathrm{d}z = \delta_{nm}. \tag{A3}$$

It can be shown by the numerical simulations that modal orthogonality is still approximately valid for the acoustic modes at two different frequencies f and f' via

$$\int_0^{\infty} \frac{1}{d(z)} u_m(f,z)u_n(f',z)\mathrm{d}z \approx \delta_{nm}, \tag{A4}$$

if the difference between f and f' is smaller than a few Hertz. Consequently, $\langle \Phi(r|r_0,f,t)\Phi^*(r|r_0,f',t')\rangle$ can be approximated as $\langle \Phi(r|r_0,f,t)\Phi^*(r|r_0,f,t')\rangle$ of Eq. (16) for sufficient narrow-band signals satisfying Eq. (A4). Equation (A2) is then approximately

$$\begin{aligned}&\langle \Psi(r|r_0,t)\Psi^*(r|r_0,t')\rangle \\ &\approx |Q(\bar{f})\Delta f|^2 \langle \Phi(r|r_0,\bar{f},t)\Phi^*(r|r_0,\bar{f},t')\rangle e^{-i2\pi\bar{f}(t-t')}, \end{aligned}\tag{A5}$$

where $\bar{f}$ and Δf are the central frequency and bandwidth of the narrow-band signal, respectively.

[1] R. J. Urick, *Sound Propagation in the Sea* (Department of Defense, Arlington, VA, 1979).
[2] R. Dashen, W. H. Munk, K. M. Watson, F. Zachariasen, and S. M. Flatte, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge, 1979).
[3] R. P. Porter, R. C. Spindel, and R. J. Jarfee, "Acoustic-internal wave interaction at long ranges in the ocean," J. Acoust. Soc. Am. **56**, 1426–1436 (1974).
[4] W. H. Munk and F. Zachariasen, "Sound propagation through a fluctuating stratified ocean: Theory and observation," J. Acoust. Soc. Am. **59**, 818–838 (1976).
[5] F. Dyson, W. Munk, and B. Zetler, "Interpretation of multipath scintillation Eleuthera to Bermuda in terms of internal waves and tides," J. Acoust. Soc. Am. **59**, 1121–1133 (1976).
[6] J. L. Spiesberger and P. F. Worcester, "Fluctuations of resolved acoustic multipaths at long range in the ocean," J. Acoust. Soc. Am. **70**, 565–576 (1981).
[7] R. P. Porter and R. C. Spindel, "Low-frequency acoustic fluctuations and internal gravity waves in the ocean," J. Acoust. Soc. Am. **60**, 943–958 (1977).
[8] Y. J. F. Desaubies, "On the scattering of sound by internal waves in the ocean," J. Acoust. Soc. Am. **64**, 1460–1469 (1978).
[9] K. D. Heaney, P. N. Mikhalevsky, H. Freese, and W. A. Kuperman, "Internal wave effects on single-frequency coherence, matched field processing, and adiabaticity," J. Acoust. Soc. Am. **103**, 2751 (1998).
[10] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Howe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, "A test of basin-scale acoustic therometry using a large aperture vertical array at 3250-km range in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[11] S. M. Flatte and G. Rovner, "Calculations of internal-wave-induced fluctuations in ocean-acoustic propagation," J. Acoust. Soc. Am. **108**, 526–534 (2000).
[12] N. C. Makris, "The effect of saturated transmission scintillation on ocean acoustic intensity measurements," J. Acoust. Soc. Am. **100**, 769–783 (1996).
[13] N. C. Makris, P. Ratilal, D. Symonds, S. Jagannathan, S. Lee, and R. Nero, "Fish population and behavior revealed by instantaneous continental-shelf-scale imaging," Science **311**, 660–663 (2006).
[14] R. N. Baer and M. D. Collins, "Source localization in the presence of internal waves," J. Acoust. Soc. Am. **118**, 3117–3121 (2005).
[15] P. Ratilal, Y. Lai, D. T. Symonds, L. A. Ruhlmann, J. R. Preston, E. K. Scheer, M. T. Garr, C. W. 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 (2004).
[16] N. C. Makris, P. Ratilal, M. Zanolin, and I. Bertsatos, "Obtaining optimal time-delay, source localization and tracking estimates in free space and in an ocean waveguide," J. Acoust. Soc. Am. **116**, 2606 (2004).
[17] K. Yoo and T. C. Yang, "Broadband source localization in shallow water in the presence of internal waves," J. Acoust. Soc. Am. **106**, 3255–3269 (1999).
[18] N. C. Makris, "Optimizing shallow-water sound-speed estimation using parameter resolution bounds," J. Acoust. Soc. Am. **96**, 3234–3235 (1994).
[19] J. L. Spiesberger, "Remote sensing of western boundary currents using acoustic tomography," J. Acoust. Soc. Am. **86**, 346–351 (1989).
[20] E. Naftali and N. C. Makris, "Necessary conditions for a maximum likelihood estimate to become asymptotically unbiased and attain the Cramer–Rao Lower Bound. Part I. General approach with an application to time-delay and Doppler shift estimation," J. Acoust. Soc. Am. **110**, 1917–1930 (2001).
[21] R. E. Williams and H. F. Battestin, "Time coherence of acoustic signals transmitted over resolved paths in the deep ocean," J. Acoust. Soc. Am. **59**, 312–328 (1976).
[22] C. Garrett and W. Munk, "Space-time scales of ocean internal waves," Geophys. Fluid Dyn. **2**, 225–264 (1972).
[23] C. Garrett and W. Munk, "Space-time scales of internal waves:a progress report," J. Geophys. Res. **80**, 291–297 (1975).
[24] P. Ratilal and N. C. Makris, "Mean and covariance of the forward field propagated through a stratified ocean waveguide with three-dimensional random inhomogeneities," J. Acoust. Soc. Am. **118**, 3532–3559 (2005).
[25] R. H. Mellen, D. G. Browning, and L. Goodman, "Diffusion loss in a stratified sound channel," J. Acoust. Soc. Am. **60**, 1053–1055 (1976).
[26] R. Dashen, S. M. Flatte, and S. A. Reynolds, "Path-integral treatment of acoustic mutual coherence functions for rays in a sound channel," J. Acoust. Soc. Am. **77**, 1716–1792 (1996).
[27] J. L. Spiesberger, F. Tappert, and A. R. Jacobson, "Blind prediction of broadband coherence time at basin scales," J. Acoust. Soc. Am. **114**, 3147–3154 (2003).
[28] N. C. Makris, F. Ingenito, and W. A. Kuperman, "Detection of a submerged object insonified by surface noise in an ocean waveguide," J. Acoust. Soc. Am. **96**, 1703–1724 (1994).
[29] F. Ingenito, "Scattering from an object in a stratified medium," J. Acoust. Soc. Am. **82**, 2051–2059 (1987).

[30]P. Ratilal, "Remote sensing of submerged objects and geomorphology in continental shelf waters with acoustic waveguide scattering," Ph.D. thesis, MIT MA, 2002.
[31]T. Chen, P. Ratilal, and N. C. Makris, "Mean and variance of the forward field propagated through three-dimensional random internal waves in a continental-shelf waveguide," J. Acoust. Soc. Am. **118**, 3560–3574 (2005).
[32]A. E. Gill, *Atmosphere-Ocean Dynamics* (Academic, San Diego, CA, 1982).
[33]P. Ratilal and N. C. Makris, "Extinction theorem for object scattering in a stratified medium," J. Acoust. Soc. Am. **110**, 2924–2945 (2001).

Passive acoustic detection of deep-diving beaked whales

Walter M. X. Zimmer[a)]
NATO Undersea Research Centre, Viale San Bartolomeo 400, 19138 La Spezia, Italy

John Harwood
Centre for Research Into Ecological and Environmental Modelling, University of St Andrews, Fife KY16 9LZ, United Kingdom

Peter L. Tyack and Mark P. Johnson
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02453

Peter T. Madsen
Zoophysiology, Department of Biological Sciences, University of Aarhus, 6000 Aarhus, Denmark

(Received 23 April 2008; revised 18 August 2008; accepted 21 August 2008)

Beaked whales can remain submerged for an hour or more and are difficult to sight when they come to the surface to breathe. Passive acoustic detection (PAD) not only complements traditional visual-based methods for detecting these species but also can be more effective because beaked whales produce clicks regularly to echolocate on prey during deep foraging dives. The effectiveness of PAD for beaked whales depends not only on the acoustic behavior and output of the animals but also on environmental conditions and the quality of the passive sonar implemented. A primary constraint on the range at which beaked whale clicks can be detected involves their high frequencies, which attenuate rapidly, resulting in limited ranges of detection, especially in adverse environmental conditions. Given current knowledge of source parameters and in good conditions, for example, with a wind speed of 2 m/s, a receiver close to the surface should be able to detect acoustically Cuvier's beaked whales with a high probability at distances up to 0.7 km, provided the listening duration exceeds the deep dive interval, about 2.5 h on average. Detection ranges beyond 4 km are unlikely and would require low ambient noise or special sound propagation conditions.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2988277]

PACS number(s): 43.30.Sf [WWA] Pages: 2823–2832

I. INTRODUCTION

Recent findings that strandings of beaked whales may be caused by human activities, such as sonar exercises (Cox *et al.*, 2006), have driven a need for improved methods to detect beaked whales. Such methods are important both for planning, so that areas with high concentrations of beaked whales can be avoided, and for mitigation to reduce the risk that sensitive animals are near ongoing exercises.

Cetaceans can be detected visually or acoustically with varying success depending on environmental conditions and their behavior (Buckland *et al.*, 2001; Barlow and Taylor, 2005). Visual detection of deep-diving cetacean species is limited in the best of sighting conditions by the short duration of their surfacing compared to their dive time. Although some species, such as sperm whales, spend considerable time visible at the surface between their deep dives and have visible blows, beaked whales only surface for a few seconds at a time and produce nearly invisible blows. The probability of visual detection for beaked whales is so low that it is of little practical use for monitoring purposes (Barlow, 1999; Barlow and Gisiner 2006). Barlow and Taylor (2005) and Barlow and Gisiner (2006) considered acoustic detection to be a more promising approach.

Whales can be detected acoustically either by using whale-finding sonar to listen for echoes reflected from the whale (active acoustic detection) or by listening for sounds emitted by the whales [passive acoustic detection (PAD)]. Whale-finding sonars do not require the whale to make a sound but to be effective, they require substantial sound energy to obtain detectable echoes. This is because the air in the lungs of a whale is compressed during a deep dive, thus reducing the whale's target strength significantly at depth. While some success in detecting baleen whales with active sonar has been reported (Lucifredi and Stein, 2007), the feasibility of active acoustic detection has not yet been demonstrated for deep-diving whales and the increased sound energy required to detect these species may generate additional risks.

PAD on the other hand is based on listening to the acoustic output from whales. Deep-diving toothed whales are known to employ echolocation to find and capture prey and are therefore acoustically very active while foraging. For example, sperm whales emit over 2500 clicks per foraging dive (Madsen *et al.*, 2002). As the spectra of sperm whale clicks reach into the audio band, they are easily detected and therefore a valuable resource for underwater bioacoustic researchers (e.g., Watkins, 1980; Gordon, 1987; Møhl *et al.*, 2000; Wahlberg, 2002; Zimmer *et al.*, 2003; Drouot *et al.*, 2004; Mellinger *et al.*, 2004; Thode, 2004; Teloni, 2005). Beaked whales also produce thousands of clicks per foraging dive

[a)]Electronic mail: zimmer@nurc.nato.int

(Johnson *et al.*, 2004, 2006; Madsen *et al.*, 2005; Zimmer *et al.*, 2005a), although most of the energy of these clicks lies above the frequencies humans can hear.

The successful application of PAD requires both appropriate technology and appropriate operational concepts. While the impact of technology or system parameters (e.g., self-noise, array gain, processing bandwidth, and gain) is easy to assess, the impact of operational factors is more difficult to quantify because they depend on the behavior of the whales (where, when, and how they echolocate) as well as on environmental characteristics that vary in time and space, influencing the design and performance of a PAD device. For example, the hydrophone depth required to maximize range and detection probability will influence the kinds of platforms that can be used to deploy it. Other operational issues include the mobility of the platform (whether the hydrophones should be moored, allowed to float, or can be towed from a ship) and the number of sensors required for obtaining the required success rate and confidence. The design of the sonar system will also depend on the objective of PAD: a system for risk mitigation, in which failure to detect a whale that is present constitutes an error, will have very different requirements than one designed for abundance estimation.

This paper outlines the requirements, and the problems, of using PAD to detect deep-diving beaked whales. In particular, we show that estimating the detection function (i.e., the probability of detecting an acoustically active deep-diving whale as function of its distance from the receiver) is a highly informative way of characterizing a PAD system.

II. METHOD AND RESULTS

Beaked whales emit two types of echolocation clicks: so-called regular clicks, suited for searching for prey items, and weaker buzz clicks used in the final stages of prey capture (Madsen *et al.*, 2005; Johnson *et al.*, 2006). In the following analysis we only consider regular clicks because they are 15 dB or more stronger than buzz clicks (Madsen *et al.*, 2002, 2005), and hence much more likely to be detected.

A. Sonar equation

The performance of a PAD system can be characterized by the passive sonar equation, which is a standardized way to assess whether or not a sound will be received with sufficient signal-to-noise ratio (SNR) for detection and classification (Lurton, 2002). The passive sonar equation combines source and environmental parameters in decibels and indicates that a sound will only be detected if the received SNR, after any filtering within the passive sonar, exceeds a predetermined receiver threshold (RT), i.e., if

$$\mathrm{SNR} > \mathrm{RT}. \tag{1}$$

Detections are not always correct because they are the result of a decision process. Incorrect detections occur when the target signal is not present but a transient in the background noise results in the received level exceeding RT. These events are called false alarms in sonar terminology or type I errors in statistics. The percentage of correct and wrong decisions is influenced by the choice of RT. The higher the value of RT, the fewer false alarms will occur, but at the same time an increasing number of weaker target signals may be missed (i.e., the proportion of type II statistical errors will increase) and the power of the detector or the probability of a correct decision will decrease.

There are a variety of approaches for selecting RT, depending on available *a priori* information and constraints. If we opt for the Neyman–Pearson criterion, RT depends on the acceptable false alarm rate P_{FA} of the detector (Lurton, 2002). That is, RT is chosen in such a way that the passive sonar system generates (false) detections in the absence of whale clicks with a predefined probability P_{FA}. If background noise is assumed to be Rayleigh distributed, RT is given by

$$\mathrm{RT} = 10\log(-2\ln(P_{\mathrm{FA}})). \tag{2}$$

In cases where the distribution of background noise is known to take a different form, RT can be derived empirically.

The detection probability P_{det} is obtained by statistical analysis of the predicted or achieved detections and depends, in general, on the relative geometry between sound source and the hydrophone of the receiver, here expressed as slant range R:

$$P_{\mathrm{det}}(R) = \Pr\{\mathrm{SNR}(R) > \mathrm{RT}\} = \int_{\mathrm{RT}}^{\infty} w(\mathrm{SNR})d\mathrm{SNR}, \tag{3}$$

where $w(\mathrm{SNR})$ is the probability density function (PDF) of the SNR at range R.

Equations (2) and (3) show that P_{det} and P_{FA} are related through RT. This relationship is often represented in the form of receiver operating characteristics (Lurton, 2002), which will also vary with the value of R. While the probability of false alarm is uniquely characterized by the noise statistics, the probability of detection depends on the properties of $w(\mathrm{SNR})$, which reflects our uncertainty about the exact received level of the whale sound.

The SNR at the detector input depends on a variety of parameters, as shown in the following equation:

$$\mathrm{SNR} = \mathrm{SL} - \mathrm{DL}(\vartheta) - \mathrm{TL}(R) - \mathrm{NL}_0 + 10\log B + \mathrm{AG} + \mathrm{PG}, \tag{4}$$

where SL is the (on-axis) source level of the sound source, $\mathrm{DL}(\vartheta)$ is the attenuation of the source level as a consequence of the off-axis angle ϑ (i.e., the angle between the acoustic axis of the source and the direction from the whale to the hydrophone),[1] $\mathrm{TL}(R)$ is the range-dependent transmission loss, NL_0 is the spectral noise level of any masking noise at the receiver position, B is the processing bandwidth, AG is the gain of the hydrophone array, and PG is the processing gain of the receiver. All of these parameters are measured in dB, except SL which is measured in $\mathrm{dB}_{\mathrm{rms}}$ re 1 μPa re 1 m,[2] NL_0 which is measured in $\mathrm{dB}_{\mathrm{rms}}$ re 1 $\mu\mathrm{Pa}^2\,\mathrm{Hz}^{-1}$, and B which is measured in Hz.

TL is a key term in the sonar equation. For relatively small values of R and constant sound speed it can be estimated by

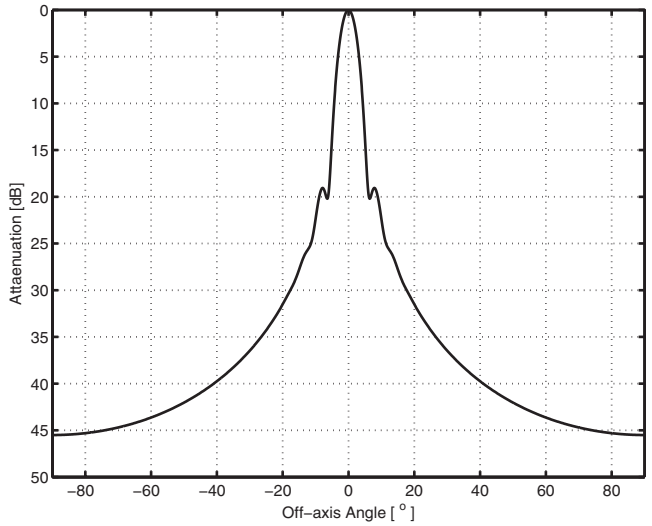


FIG. 1. Off-axis dependent attenuation of the echolocation signal of Cuvier's beaked whale adapted from Fig. 7 of Zimmer *et al.* (2005b).

$$\mathrm{TL}(R) = 20\log(R) + \alpha_f(R/1000), \tag{5}$$

where range R is expressed in m and α_f is the frequency dependent absorption coefficient in dB km^{-1}.

Clicks of Cuvier's and Blainville's beaked whales are centered around 40 kHz (Johnson *et al.* 2004, 2006; Zimmer *et al.*, 2005a), for which the absorption coefficient is about 9.5 dB km^{-1} (Lurton, 2002).

We assume that local ambient noise is mainly generated by surface waves and can therefore be described by the Knudsen curves [Lurton, 2002, Eqs. (4.12)]. This noise will decrease as it propagates from the surface to the bottom. Thus the SNR, and therefore P_{det}, will increase with the receiver depth [Lurton, 2002, Eq. (4.41)].

While typical values can be used for most of the quantities in the passive sonar equation, the off-axis dependent attenuation of the source level, or off-axis loss, DL(ϑ) of echolocating whales needs to be evaluated explicitly. Here, we consider that beaked whales have a broadband biosonar system (Zimmer *et al.*, 2005a) and that we have a rough knowledge of the transmission beam pattern, as presented in Fig. 1. The beam pattern behind the animal ($|\vartheta| > 90°$) is assumed constant at the maximum attenuation level.

B. Basic detection function for a single click

A basic detection function for a single beaked whale click can be determined from Eq. (3), provided that suitable values for the constants of the sonar equation are available to estimate the SNR [Eq. (4)]. In order to illustrate this process we have used the values shown in Table I. In Sec. III we describe the sensitivity of our results to these values. The spectral noise level in Table I corresponds to a light breeze or a wind speed of 2.5 m s^{-1} for a near surface hydrophone, and the receiver threshold of 14 dB is for Rayleigh distributed ambient noise equivalent to $P_{\mathrm{FA}} = 3.5 \times 10^{-6}$. This is equivalent to accepting a false detection about every minute for a processing window of 0.2 ms.

The relationship between SNR, R, and ϑ is shown in Fig. 2; the white threshold line divides the SNR map into two parts: to the left of this line, the SNR exceeds RT and detection is possible, but to the right the SNR is too low for a successful detection.

TABLE I. Constants of the sonar equation used to construct Figs. 2 and 3. Values were chosen to be characteristic of beaked whales and are consistent with Johnson *et al.* (2004) and Zimmer *et al.* (2005a). Effects of depth on the spectral noise level are not included.

Source level	200 dB$_{\mathrm{rms}}$ re 1 μPa at 1 m
Center frequency	40 kHz
Processing bandwidth	40 kHz
Spectral noise level	30 dB re 1 μPa^2/Hz
Array gain	0 dB
Processing gain	0 dB
Sound attenuation	9.5 dB/km
Receiver threshold	14 dB

Figure 2 was constructed on the assumption that nothing is known about the distribution of the aspect of an echolocating beaked whale with respect to the receiver. Nevertheless, one can deduce some important results from Fig. 2. In particular, for the values in Table I, a single beaked whale is very unlikely to be detected at ranges greater than 4 km because the SNR for all off-axis angles is below the threshold for the assumed propagation condition and source parameters. Using similar arguments, we can conclude that all clicks will be detected for ranges below about 0.7 km as the SNR for all off-axis angles is above the threshold. Between these limits, detectability depends on the probability that the whale is clicking in the direction of the receiver. Specific detection probabilities can be obtained if the orientation statistics for a particular location of animal and receiver are available and all other terms of SNR [Eq. (4)] are known because Eq. (1) is then equivalent to

$$P_{\mathrm{det}}(R_H, h) = \int_{\mathrm{SNR}(\vartheta) > \mathrm{RT}} w(\vartheta)\, d\vartheta, \tag{6}$$

where R_H is the horizontal displacement of the whale from the hydrophone, h is the depth of the hydrophone, and $w(\vartheta)$

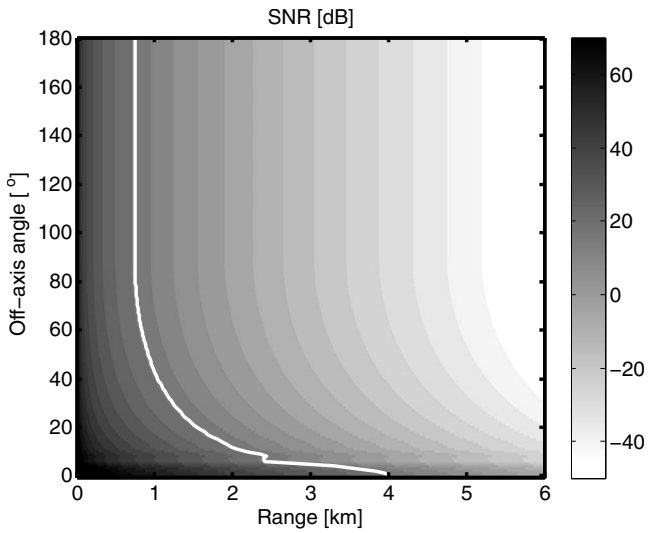


FIG. 2. SNR as a function of the off-axis angle of the source and its slant range. The white line indicates the threshold RT=14 dB.

is the PDF of the off-axis angle and the integral runs over all off-axis angles for which the SNR exceeds RT.

C. Off-axis angle statistics

The actual off-axis angle depends on the location of the whale relative to the hydrophone and its orientation while echolocating. If we assume that the whale orientation is not correlated with the location of the hydrophone, then we may choose any coordinate system, for example, [E, N, Z], where the x-axis corresponds to east (E), the y-axis to north (N), and the z-axis to the zenith (Z). The acoustic axis or direction vector of the whale when making a particular click is given by

$$\hat{c} = \begin{pmatrix} c_E \\ c_N \\ c_Z \end{pmatrix} = \begin{pmatrix} \sin\gamma\cos\beta \\ \cos\gamma\cos\beta \\ \sin\beta \end{pmatrix}, \tag{7}$$

where γ is the whale's heading (measured from north to east) and β represents its pitch (positive pitch is measured upward). The direction of the hydrophone, as seen by the whale, is

$$\hat{s} = \begin{pmatrix} s_E \\ s_N \\ s_Z \end{pmatrix} = \begin{pmatrix} \cos\eta \\ 0 \\ \sin\eta \end{pmatrix}, \tag{8}$$

where η is the elevation angle of the hydrophone from the whale, which is given by

$$\eta = \tan^{-1}\left(\frac{d-h}{R_H}\right), \tag{9}$$

where d is the depth of the whale and h is the depth of the hydrophone. In Eq. (8), the hydrophone is modeled as being due east of the whale, but there is no loss of generality as the whale can be in an arbitrary orientation via Eq. (7).

The off-axis angle ϑ (i.e., the angle between the whale's acoustic axis and the hydrophone direction) is then given by

$$\cos\vartheta = \hat{c}^T\hat{s}, \tag{10}$$

that is,

$$\vartheta(\gamma,\beta,\eta) = \cos^{-1}(\sin\gamma\cos\beta\cos\eta + \sin\beta\sin\eta), \tag{11}$$

and slant range R is simply determined from the relationship

$$R = \sqrt{R_H^2 + (d-h)^2}. \tag{12}$$

The way in which variations in off-axis angle affect the form of the detection function will depend on how a detection is defined. Here we consider a situation in which we want to determine the probability that at least one click will be detected during an entire dive sequence. In this case, all we need to know is the minimum off-axis angle adopted by a whale within a dive:

$$\vartheta_{\min}(\beta,\eta) = \min_{\text{dive}}\{\cos^{-1}(\sin\gamma\cos\beta\cos\eta + \sin\beta\sin\eta)\}. \tag{13}$$

If the SNR associated with this orientation does not exceed RT, then the whale cannot be detected. If the SNR does exceed RT, then other clicks within the dive may also be detected, but this will not affect the dive P_{det}. Ideally, statistical distributions for γ, β, and η should be obtained from direct observations of beaked whale dives. However, although some data for beaked whales are available (Tyack *et al.*, 2006; Johnson *et al.*, 2006), these are limited to a very small sample of animals. Instead, we assumed distributions for the parameters that would be expected for a beaked whale foraging on prey that are uniformly distributed in a horizontal layer around the predator. We recognize that beaked whales may feed on more than one layer, and in Sec. III we consider the sensitivity of our results to this simplification.

The distribution of whale pitch was modeled as a circular normal, or von Mises, distribution (Fisher, 1993):

$$w(\beta) = \frac{1}{2\pi I_0(\kappa)}\exp(\kappa\cos(\beta)), \tag{14}$$

where κ is a parameter related to the variance of the distribution by $\text{var}(\beta) = 1 - I_1(\kappa)^2/I_0(\kappa)^2$ and $I_j(x)$ is the modified Bessel function of order j. For the present analysis a value of $\kappa = 6$ was chosen, equivalent to a standard deviation of about 23°, which means that only a few clicks are oriented straight up or down, consistent with foraging in a horizontal layer. The distribution of heading within a dive was also modeled as a circular normal distribution relative to a mean whale heading γ_0 with the same standard deviation as for the pitch distribution, i.e., $\kappa = 6$.

The distribution of the hydrophone elevation relative to the whale was modeled as a function of hydrophone depth and horizontal distance according to Eq. (9). We assumed that whale depth during a dive was normally distributed around a mean foraging depth d_0.

$$w(d) = \frac{1}{\sqrt{2\pi}\sigma_d}\exp\left(-0.5\left(\frac{d-d_0}{\sigma_d}\right)^2\right). \tag{15}$$

For the present analysis d_0 was set to 720 m, with a standard deviation $\sigma_d = 50$ m, and h to 100 m. Later, we examine the consequences of choosing different values for these parameters. We considered horizontal ranges from 0 to 10 km.

To obtain the PDF of the off-axis angle for a given scenario, or geometry, the assumed distributions $w(\gamma)$, $w(\beta)$, and $w(\eta)$ were transformed into $w(\vartheta)$. Although this could have been done analytically, it was straightforward to use a Monte Carlo simulation. In order to simulate a single dive, we drew 4000 random values from each of the distributions defined above. These were then transformed into a vector of random off-axis angles. This vector was used to simulate a whale's orientation each time it clicked during that dive. Although the pitch, heading, and elevation of a whale when a click is generated are likely to be highly correlated with its orientation on previous and subsequent clicks, our concern is only whether or not any of the clicks made during a dive can be detected. In this circumstance, the assumption of independence for the three PDFs is justified.

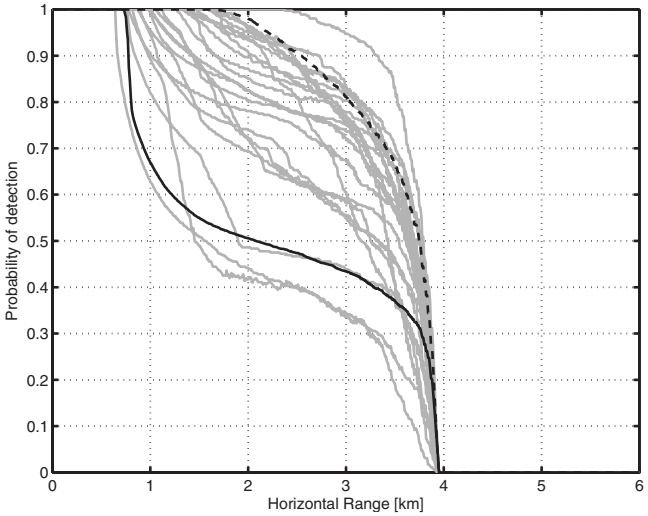


FIG. 3. Detection functions for a beaked whale echolocating at a depth of 720 m with a receiver depth of 100 m. The solid black curve corresponds to a simple simulation with normally distributed heading, pitch, and elevation. The gray lines show the probability of detection for clicks generated during 23 measured dives made by six Cuvier's beaked whale (Zc). The dashed black curve shows the effect of allowing a simulated whale to reverse the direction of its travel periodically.

D. Detection function for an individual dive

The process of dive simulation was repeated for all mean headings relative to the hydrophone (i.e., we rotated the mean whale heading γ_0 around the circle in steps of 0.1°) to obtain a range-dependent PDF for the off-axis angle. The detection function was then estimated by integrating this PDF according to Eq. (6). The result of this integration is shown by the solid black curve in Fig. 3.

This detection function has some properties in common with the one shown in Fig. 2. For example, all dives made within 0.7 km horizontal separation of the hydrophone from the whale are certain to be detected and no dives made beyond 4.0 km will be detected, for the source and environmental parameters used here, because absorption limits the detectability even of on-axis clicks. However, the distribution of detection probabilities within this range will vary, depending on the actual movements of the whale.

For comparison, we have also included 23 detection functions (shown in gray) for real dive profiles for six Cuvier's beaked whales (*Ziphius cavirostris*) in Fig. 3. For this data set the random pitch, heading, and depth distributions were replaced by actual measurements collected by digital acoustic recording tags (DTAG; Johnson and Tyack 2003) in the Ligurian Sea during 2003 and 2004 (Tyack *et al.*, 2006), that is, the detection functions were estimated in analogy to the simulated dives with the exception that whale pitch, heading, and depth were not modeled according to Eqs. (14) and (15) but obtained from the DTAG. The dashed black line in Fig. 3 was obtained by replacing the monomodal heading PDF [Eq. (14)], as used for the solid black line in Fig. 3, by a dual-modal PDF simulating a reversal of the swimming direction with a probability of 50%. This would represent a feeding behavior of directional swimming back and forth, for example, along a certain bathymetric contour. Both the solid and the dashed line in Fig. 3 characterize the boundaries of the simulated detection functions. As seen from Fig. 3, most detection functions based on DTAG data fall within these limits and the few exceptions (e.g., to the left of the solid line) are due to residual mismatch in modeling parameters, for example, deeper foraging depth would result in shorter detection ranges.

III. DISCUSSION

In this paper, we have used diving and acoustic data of beaked whales along with theoretical considerations based on the sonar equation to explore the potential and limitations of PAD for deep-diving echolocating beaked whales. While the sonar equation is the tool of choice for assessing the performance of human sonar systems, it is a simplified mathematical description of a complicated problem. We simplified this system even further by analyzing only a single, but characteristic, parameter set shown in Table I. However, the sensitivity of our results to variation in these parameters can, in most cases, be evaluated in a straightforward way.

A. Importance of the components of the sonar equation

Many of the components of the sonar equation have a simple additive effect on the SNR associated with a whale click and therefore on the probability that it will be detected. Thus the effect of a 10 dB reduction in source level, a 10 dB increase in ambient noise at the depth of the PAD hydrophone, or a 10 dB decrease in the gain of the hydrophone array or the receiver will all have an identical effect on the detection function. The consequence of this 10 dB change is to reduce the distance over which whales are certain to be detected from 0.7 to 0.3 km and to reduce the maximum distance at which whales can be detected from 4 to 3.1 km. Similarly, an increase in the SNR by 10 dB would increase the lower limit of the detection function from 0.7 to 1.3 km, and the upper limit from 4.0 to 4.9 km. The shape of the detection function between these two end points remains nearly unchanged.

The most important component in the sonar equation is the on-axis source level of the clicking whale for which a nominal value of 200 dB_{rms} re 1 μPa at 1 m was assumed in constructing Fig. 3. However, the source level of regular clicks during deep dives is not constant. For example, the source level of sperm whale echolocation clicks is known to vary by up to 20 dB (Madsen *et al.*, 2002) and the apparent output of beaked whale echolocation clicks seems to vary by a similar factor (Madsen *et al.*, 2005). However, the effect of SL on SNR is strictly additive. Therefore, the consequence of a statistical variation in SL will be to decrease the distance at which whales are certain to be detected and to increase the maximum detection distance.

Altering the RT has a similar effect to changing SL, but on the probability of detection axis of Fig. 3. Increasing RT will decrease the distance at which whales can be detected with certainty (possibly to zero) and decrease the maximum distance at which whales can be detected but will also reduce the rate of false alarms.

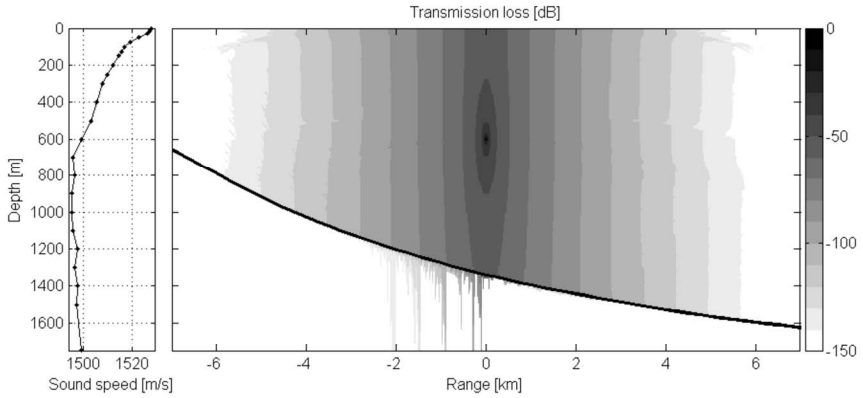


FIG. 4. TL estimation using an average sound-speed profile (left panel). The TL (right panel) was estimated using Bellhop (Porter and Bucker, 1987) for a 40 kHz omnidirectional signal with an assumed bottom profile as shown.

The ambient noise level, which masks the whale sound during reception, is the second most important parameter in the sonar equation. Ambient noise is very difficult to model. Most textbook descriptions of sonar performance base their high-frequency noise level estimation on the Knudsen model for surface-generated noise (Urick, 1983; Burdic, 1984; Lurton, 2002). Actual noise levels are very often higher than predicted by the Knudsen model due to additional noise sources such as rain, thunderstorms, and shipping (Wenz curves; Urick, 1983; Aguilar *et al.*, 2006). It is essential to measure the ambient noise level *in situ* at the frequencies of interest to predict its impact on the sonar performance in any particular site.

Surface-generated noise decreases as a function of depth. At the frequencies used by beaked whales for echolocation, surface-generated ambient noise is reduced by more than 20 dB for a hydrophone depth of 2000 m [Lurton, 2002, Eq. (4.41)]. This means that deep hydrophones, in general, will perform better than shallow ones provided they are no more than 1 km below the whale foraging depth. Deep hydrophones will also be more likely to detect the downward-directed clicks produced during descent. However, operational constraints (see below) may limit the depth at which hydrophones can be deployed. The impact of ambient noise may also be reduced by increasing the gain of the array (AG) and/or the receiver (PG). Increasing AG requires increased investment in hardware (more hydrophones, more processing power, etc.), thus reducing the cost effectiveness of PAD.

PG is related to the analysis bandwidth, which is typically adapted to the type of signal to be detected. The maximum PG can be estimated based on the signal characteristics of beaked whale clicks. Both species analyzed so far (Cuvier's and Blainville's beaked whales) show a unique time-variable spectrum of short frequency-modulated upsweeps (Zimmer *et al.*, 2005a; Johnson *et al.*, 2006), at least when recorded close to the acoustic axis. To take advantage of this feature, a PAD processor could implement a matched filter to emphasize beaked whale signals resulting in a PG of 10 log (BT), where BT is the time-bandwidth product. For Cuvier's beaked whale, this would be about 9 dB (signal bandwidth of 40 kHz and signal length of 0.2 ms). This gain may be reduced when echolocation clicks are detected at large off-axis angles because frequency dependent distortions would then become significant. As a result, the performance of the matched filter would decrease due to increasing mismatch between the received signal and its replica in the filter.

In our calculations, TL was modeled using the spherical spreading law. Strictly speaking, this is only valid in an unbounded medium with constant sound speed. Such theoretical conditions never exist in real oceans, especially for long distances, where the presence of ocean boundaries, refraction, and scattering results in the need for complicated models of TL. Nevertheless, the spherical spreading law is a reasonable first approach in assessing the performance of PAD over the relatively short distances in question here. However, it should be recognized that performance at longer ranges may increase due to sound trapping or decrease due to shadow zones (e.g., see Lurton, 2002, Fig. 2.25 for propagation in deep sound channel and Fig. 2.27 for shadow and convergence zones). Prediction of TL for any specific site requires acoustic propagation modeling that includes realistic sound-speed profiles in combination with a detailed description of the bottom characteristics of the site. Figure 4 shows as example a TL estimation for a late summer sound-speed profile typically found in subtropical Atlantic climate (28° latitude). The Bellhop model (Porter and Bucker, 1987) predicts for a beaked whale foraging at 600 m depth, a TL of about 130 dB at a horizontal distance of 5 km.

The final relevant component of the sonar equation is the attenuation resulting from off-axis transmission of echolocation clicks. The sonar of echolocating whales, measured to date, is characterized by rather narrow sound beams with directivity indices normally more than 25 dB (Au, 1983; Zimmer *et al.*, 2005a, 2005b). This is equivalent to a half-power beam width of less than 10° (Zimmer *et al.*, 2005a) and results in a substantial decrease in apparent source level at large off-axis angles. For Cuvier's beaked whale, the maximum off-axis attenuation has been measured to be about 46 dB (Zimmer *et al.*, 2005a). This narrow beam width translates to a low probability that a random single click would be detected at a distance by a single PAD system, as shown clearly in Fig. 2. Fortunately, echolocating whales emit a large number of clicks in varying directions during each foraging dive. This substantially increases the probability that a given PAD system will detect at least one click, as shown by our modeling of the consequence of the variations in orientation during the foraging portion of a dive (Fig. 3).

The depth at which beaked whales forage (d), and therefore produce clicks consistently, is likely to vary both within and, particularly, between dives, and from location to location. Dive depth will affect both the slant range to the hydrophone of the PAD system [Eq. (12)] and the elevation of the hydrophone from the whale [η, Eq. (9)]. However, these effects are greatest at distances of less than 1 km (i.e., in the region where detection of at least one click is certain). At distances of greater than 1 km, the effect of a $\pm 10\%$ variation in d is relatively small ($\pm 1°$ for η and $\pm 10\%$ for R) provided the difference between whale and hydrophone depth is not too great. We therefore argue that the shape of the detection function is not particularly sensitive to variations in whale depth.

B. Detection functions

The maximal detection range (4 km in Figs. 2 and 3) and the minimal range, within which no echolocation clicks are missed (0.7 km), depend only on the two extremes of the off-axis attenuation and are independent of the actual distribution of off-axis angles. The maximal range where whales may be detected corresponds to the limit for the detection of on-axis clicks, and the guaranteed detection range, where echolocation clicks are always detectable, results from clicks observed with the maximum off-axis attenuation.

The actual shape of the detection function depends on the precise form of the PDF for off-axis angles. Here we have modeled this distribution for a single complete dive, containing 4000 clicks, so that the resulting detection function indicates the probability of detecting at least one click during the dive. The resulting detection function indicates that about 25% of dives produce clicks that are close enough to on axis to be detected at a range of up to 3.8 km. A further 25% of the dives produced clicks so far off axis that they can only be detected when the animal is within 0.8 km of the hydrophone.

We compared the simulated detection function (solid black line in Fig. 3) with a detection function derived from measured values for the depth, pitch, and heading for each click made by real Cuvier's beaked whales during the course of foraging dives (gray lines in Fig. 3). When we examined the distribution of headings for these animals, we found that it did not correspond very closely to that used to generate the simulation (solid black line in Fig. 3). In particular, some of the whales tended to reverse heading by about 180°, as might be expected if they were carrying out an area-restricted search through a prey concentration. This behavior resulted in a PDF of heading that had two modes. The circular normal distribution does not account for such inversions of direction. We therefore developed another kind of simulated dive during which 50% of the clicks had a mean heading γ_i and the remaining 50% had a mean heading ($\gamma_i + 180°$). The resulting detection function is shown by the dashed black curve in Fig. 3.

The two simulated detection functions (with and without reversal of swimming direction (solid and dashed lines, respectively, in Fig. 3) bracket the empirical detection functions derived from direct measurements of beaked whale dives, suggesting that our method provides a realistic model for estimating general detection functions.

Beaked whales initiate regular clicking during the descent phase of a deep dive, at a depth of about 400–500 m, usually well above the foraging layer (Tyack *et al.*, 2006) whereby these clicks are usually downward oriented and are therefore likely to be detected only by hydrophones below the mean foraging depth or at very short ranges by hydrophones located close to the surface, as indicated for large off-axis angles in Fig. 2.

C. Classification of detected clicks

Any real detector produces false alarms and in practical applications, it is necessary to evaluate each detection to assess the likelihood that it represents a real signal. Detecting beaked whales then involves a two-step process: detection of clicklike sounds and classification of these sounds as stemming from a beaked whale. Classification performance can be enhanced not only by considering the spectral features of individual clicks, as described above, but also by analyzing the temporal characteristics of click sequences, taking advantage of the observation that echolocating animals tend to produce sequences of clicks with a slowly varying interclick interval (ICI).

Temporal classification requires the detection of consecutive clicks to estimate ICI. As Fig. 1 indicates, the acoustic beam produced by beaked whales is rather narrow (here modeled with a 6° beamwidth at −3 dB). As a result, a receiver is likely to only record a series of clicks (known as a "scan") over a short time interval. Clicks within scans can vary in level by more than 20 dB as the beam crosses the hydrophone location (Johnson *et al.* 2006, Madsen and Wahlberg, 2007). The objective of detecting at least one click made by an echolocating whale can easily be transformed to one of detecting at least one scan if RT is increased by 20–30 dB. Different odontocete species have differing but overlapping distributions of ICIs. Cuvier's beaked whales have a mean ICI of 0.4 s (Zimmer *et al.*, 2005a) whereas sperm whales tend to use ICIs > 0.5 s (Zimmer *et al.*, 2005b). Dolphins typically have an ICI of less than 0.1 s (Madsen *et al.*, 2004) and so might be relatively easily differentiated from beaked whales on this basis.

Single click classification is, in general, spectrally based and requires a good SNR over an extended signal band to be successful. The spectral description of Cuvier's beaked whale click (Zimmer *et al.*, 2005a) indicates that spectral energy 10 dB below the peak signal value may be sufficient to differentiate a single click from those of other nonziphiid echolocating toothed whales. Increasing RT by 10 dB would ensure that only signals that allow spectral classification are detected.

An advanced detector should be able to adapt to the environment and change its parameter settings (e.g., detection threshold and analysis bandwidth) according to the success or failure of past decisions. Wald (1947) described a decision scheme based on a sequential probability ratio test, which does not require every detection opportunity (e.g., ev-

ery detected transient) to be classified. If it is uncertain, the detector may defer a decision until after the reception of a new data set. This scheme is repeated until the desired decision (beaked whale present or absent) can be made. Such a scheme would initially try to classify the detection according to spectral features but, if uncertain, would change its classification strategy as the number of clicks with the "correct" ICI in a sequence increases.

The relatively low detection threshold (RT=14 dB) used here will give a high false alarm rate, making classification essential for most applications. However, the addition of a classifier will modify the overall detection function and it is important to re-evaluate the detection function for the entire PAD system once the detection-classification algorithms are established.

As discussed above, the performance of a PAD hydrophone should improve as its depth is increased because of the lower ambient noise. This suggests that hydrophones mounted on, or close to, the bottom, such as those at the AUTEC naval underwater range (Moretti *et al.*, 2006), should outperform hydrophones close to the surface. However, operational considerations may favor shallower hydrophone deployments. For example, it is difficult to deploy a towed hydrophone at or below the mean foraging depth of beaked whales. An additional advantage of a hydrophone that is situated above the whale is that it may be able to exploit surface reflections of upward-directed clicks to provide an additional classification cue. The delay between the direct arrival of a click and its surface reflection depends only on the geometry between whale and the hydrophone, which should remain fairly constant during the short duration of a click sequence. The resulting stable double click could therefore be a useful cue for classification of clicks stemming from an individual deep-diving whale. The decreased SNR of a shallow hydrophone may therefore be compensated by an improved classification performance, which would allow a reduction in RT.

D. Complicating factors

The above analyses assume that clicks are emitted by a single animal. It is therefore important to understand what will happen when more than one animal is present. A naïve suggestion would be that detection probability scales directly with the number of clicking animals. However, the level of improvement will depend on the extent to which the animals' orientations are correlated and on the performance of the receiver in the presence of multiple animals.

Surface observations indicate that members of beaked whale groups tend to surface and dive in close vicinity to each other but little is known about subsurface behavior of groups of whales, especially with respect to their interindividual relative orientation. Zimmer *et al.* (2005a) showed that a pair of Cuvier's beaked whales remained within a few hundred meters of each other during a foraging dive. As a result, the cumulative acoustic footprint of a group, i.e., the acoustically illuminated volume, and hence the probability that it will be detected, may change little with group size.

The presence of multiple foraging animals may also confound the classification process, especially when classification is based on ICI. However, the length of time that a single whale directs detectable clicks toward the hydrophone is likely to be limited to a few seconds at a time due to the narrow sonar beam and the constant movement of the animals. Consequently, provided that individual whales orient independently, their clicks will be received by a hydrophone at different times, thus reducing the interference. Within-scan interference by conspecifics can be eliminated by increasing the RT so that only high-level on-axis clicks are used for classification purposes. Nonetheless, the potential difficulties associated with classification based on temporal cues underline the need for classification concepts that are based on spectral or joint information.

In addition to conspecific interference, there is also the potential of interference from other cetacean species. For example, dolphins also emit short echolocation clicks that are characterized by a very broadband spectrum that, in most cases, overlaps with the spectrum of clicks emitted by beaked whales. However, discrimination of dolphins from beaked whales may be aided by the presence of whistles, their shorter ICIs, and the lack of significant click energy below 20 kHz in beaked whale clicks (Zimmer *et al.*, 2005a, Johnson *et al.*, 2006).

E. Operational considerations

The way in which a PAD system is operated will influence detection performance. A key operational parameter is how long the PAD system should monitor a water volume before making a decision about the presence or absence of the target species. This depends on the specific purpose of the whale detector. If the main aim is the mitigation of anthropogenic impacts on deep-diving whales, then an obvious sampling interval is one that is long enough to include at least one foraging dive and the interval between such dives.

Foraging beaked whales echolocate in regular bouts of duration about 30 min for Cuvier's beaked whales (Tyack *et al.*, 2006), but the time between consecutive deep dives, that is the time where whales are silent, is variable and can exceed 110 min. This yields an initial requirement for a listening time of at least 140 min in order to have a high probability of detecting Cuvier's beaked whales. The optimal listening time may be longer both to sample multiple dive cycles and because our estimates of silent durations are drawn from a limited data set and therefore may be biased toward short intervals.

If PAD is to be used as part of a survey whose purpose is to estimate whale density, then it is desirable to estimate the distance of the whale from the hydrophone when it is detected. This can be done either by detecting the same whale simultaneously with multiple PAD systems and determining its location by triangulation or by exploiting information from the multipath structure of sound propagation, if this is available. For example, a compact three-dimensional array of hydrophones may estimate the direction to the whale and can make use of a single surface reflection to estimate the range and depth of a clicking whale.

All of the calculations presented here have been based on the assumption that the PAD hydrophone is stationary. If this assumption is relaxed, the assessment of PAD performance becomes more complicated and requires detailed modeling of the relative motion of the whale and hydrophone to determine a dynamic acoustic detection function. The acceptable platform speed will depend on the goals of PAD. For risk mitigation requiring a high level of confidence that the absence of detections was indeed due to the absence of whales, relatively low PAD platform speeds, probably less than 4 km h^{-1}, are needed to ensure adequate time in each water volume. These speeds may be best achieved by mounting hydrophones on deep-diving gliders or drifting buoys. On the other hand, if PAD is being used to determine whale density, it may be more cost effective to tow a hydrophone behind a vessel. The reduced probability of detecting whales when they are present would be compensated by the ability to survey a larger area within a given time or for a given cost.

IV. CONCLUSIONS

At present, PAD offers the only reliable method for determining the presence or absence of deep-diving echolocating beaked whales, which are difficult, and often impossible, to detect visually. Here we show that the performance of a PAD system can be characterized using a detection function, which describes how the probability of detecting an acoustically active whale varies as function of its distance from the receiver, derived using the sonar equation. This indicates that a beaked whale is almost certain to be detected during a foraging dive if it is within approximately 0.7 km of a PAD system and unlikely to be detected if it is more than 4 km away for the assumed propagation conditions and click source properties. The precise limits for detection depend on the source level of the whale's clicks, the ambient sound level at the depth of the hydrophone, and the gain settings of the hydrophone arrays and receiver. Between these limits, the form of the detection function depends on the way in which the orientation of the whale varies with respect to the PAD hydrophone during the course of a dive. We show, by comparison with detection functions derived using data collected directly from diving whales, that these variations can be represented by a simple statistical model.

Some of the factors that affect the performance of PAD (the quality and quantity of the listening equipment) are under the control of the observer. However, the whales control their sound emissions and the environment determines the propagation and the background noise level, which masks the whale's echolocation clicks. Because the performance of PAD depends heavily on environmental conditions, it must be implemented carefully to compensate, at least partially, for the effect of these conditions.

The availability of beaked whales for acoustic detection depends on the interval between foraging dives. This silent interval may exceed 110 min so that extended listening periods will be required to ensure a high probability that whales are detected by a PAD system. This suggests that if a PAD system is being used to mitigate the risks to beaked whales associated with anthropogenic sound, it will be most effective if its hydrophones are mounted on slow moving platforms (such as gliders or drifting buoys) rather than a fast moving active sonar vessel.

The detection functions presented here are based on a single choice of parameters. Empirical detection functions could and should be used to calibrate the different model assumptions. As more dive data are measured from beaked whales, statistical analysis of relevant dive parameters may be used to improve the associated PDF. As different applications (risk mitigation or abundance estimation) have different requirements, it will also be important to analyze the sensitivity of their success to the actual shape of the detection function.

ACKNOWLEDGMENTS

This work was carried out within the Marine Mammal Risk Mitigation Project of the NATO Undersea Research Center. We thank Len Thomas and Tiago Marques for comments and suggestions. Cuvier's beaked whale data were collected by the Woods Hole Oceanographic Institution and BluWest S.R.L. with funding from SERDP and the National Oceanographic Partnership Program (NOPP). P.T.M. was funded by a Steno grant from the National Danish Science Foundation. M.P.J. and P.L.T. were funded under the Office of Naval Research, SERDP and the NOPP. 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.

[1]An off-axis angle $\vartheta=0$ is "on axis."

[2]But see Urick (1962) for a general sonar equation suited also for transients; see also Au (1993).

Aguilar de Soto, N., Johnson, M., Madsen, P. T., Tyack, P. L., Bocconcelli, A., and Borsani, J. F. (**2006**). "Does intense ship noise disrupt foraging in deep-diving Cuvier's beaked whales (*Ziphius cavirostris*)?" Marine Mammal Sci. **22**, 690–699.

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

Barlow, J. (**1999**). "Trackline detection probability for long-diving whales," in *Marine Mammal Survey and Assessment Methods*, edited by G. W. Garner, S. C. Amstrup, J. L. Laake, B. F. J. Manly, L. L. McDonald, and D. G. Robertson (A.A. Balkema, Rotterdam), pp. 209–224.

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. L. (**2005**). "Estimates of sperm whale abundance in the northeastern temperate Pacific from combined acoustic and visual survey," Marine Mammal Sci. **21**, 429–445.

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, UK).

Burdic, W. S. (**1984**). *Underwater Acoustic System Analysis* (Prentice-Hall, Englewood Cliffs, NJ).

Cox, T. M., Ragen, T. J., Read, A. J., Vos, E., Baird, R. W., Balcomb, K., Barlow, J., Caldwell, J., Cranford, T., Crum, L., D'Amico, A., D'Spain, G., Fernández, A., Finneran, J., Gentry, R., Gerth, W., Gulland, F., Hildebrand, J., Houser, D., Hullar, T., Jepson, P. D., Ketten, D., MacLeod, C. D., Miller, P., Moore, S., Mountain, D. C., Palka, D., Ponganis, P., Rommel, S., Rowles, T., Taylor, B., Tyack, P., Wartzok, D., Gisiner, R., Mead, J., and Benner, L. (**2006**). "Understanding the impacts of anthropogenic sound on beaked whales," J. Cetacean Res. Manage. **7**, 177–187.

Drouot, V., Gannier, A., and Goold, J. C. (**2004**). "Summer social distribu-

tion of sperm whales (*Physeter macrocephalus*) in the Mediterranean Sea," J. Mar. Biol. Assoc. U.K. **84**, 675–680.
Fisher, N. I. (**1993**). *Statistical Analysis of Circular Data* (Cambridge University Press, Cambridge, UK).
Gordon, J. C. D. (**1987**). "The behaviour and ecology of sperm whales off Sri Lanka," Ph.D. thesis, Darwin College, Cambridge.
Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N., and Tyack, P. L. (**2004**). "Beaked whales echolocate on prey," Proc. R. Soc. London, Ser. B **271**, S383–S386.
Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N., 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.
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.
Lucifredi, I., and Stein, P. J. (**2007**). "Gray whale target strength measurements and the analysis of the backscattered response," J. Acoust. Soc. Am. **121**, 1383–1391.
Lurton, X. (**2002**). *An Introduction to Underwater Acoustics* (Springer, London).
Madsen, P. T., Johnson, M., Aguilar de Soto, N., Zimmer, W. M. X., and Tyack, P. L. (**2005**). "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)," J. Exp. Biol. **208**, 181–194.
Madsen, P. T., Kerr, I., and Payne, R. (**2004**). "Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: False killer whales (*Pseudorca crasidens*) and Risso's dolphins (*Grampus griseus*)," J. Exp. Biol. **207**, 1811–1823.
Madsen, P. T., and Wahlberg, M. (**2007**). "Recording and quantitative analysis of clicks from echolocating toothed whales in the wild," Deep-Sea Res., Part II **54**, 1421–1444.
Madsen, P. T., Wahlberg, M., and Møhl, B. (**2002**). "Male sperm whale (*Physeter macrocephalus*) acoustics in a high-latitude habitat: Implications for echolocation and communication," Behav. Ecol. Sociobiol. **53**, 31–41.
Mellinger, D. K., Stafford, K. M., and Fox, C. G. (**2004**). "Seasonal occurrence of sperm whale (*Physeter macrocephalus*) sounds in the gulf of Alaska, 1999–2001," Marine Mammal Sci. **20**, 48–62.
Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). "Sperm whale clicks: Directionality and source level revisited," J. Acoust. Soc. Am. **107**, 638–648.
Moretti, D., Morrissey, R., DiMarzio, N., and Ward, J. (**2006**). "Verified passive acoustic detection of beaked whales (*Mesoplodon densirostris*) using distributed hydrophones in the tongue of the ocean, Bahamas (A)," J. Acoust. Soc. Am. **119**, 3374.
Porter, M. B., and Bucker, H. P. (**1987**). "Gaussian beam tracing for computing ocean acoustic fields," J. Acoust. Soc. Am. **82**, 1349–1359.
Teloni, V. (**2005**). "Patterns of sound production in diving sperm whales in the northwestern Mediterranean," Marine Mammal Sci. **21**, 446–457.
Thode, A. (**2004**). "Tracking sperm whale (*Physeter macrocephalus*) dive profiles using a towed passive acoustic array," J. Acoust. Soc. Am. **116**, 245–253.
Tyack, P. L., Johnson, M., Soto, N. A., Sturlese, A., and Madsen, P. T. (**2006**). "Extreme diving of beaked whales," J. Exp. Biol. **209**, 4238–4253.
Urick, R. J. (**1962**). "Generalized form of the sonar equations," J. Acoust. Soc. Am. **34**, 547–550.
Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).
Wahlberg, M. (**2002**). "The acoustic behaviour of diving sperm whales observed with a hydrophone array," J. Exp. Mar. Biol. Ecol. **281**, 53–62.
Wald, A. (**1947**). *Sequential Analysis* (Wiley, New York).
Watkins, W. (**1980**). "Acoustics and the behavior of sperm whales," in *Animal Sonar Systems*, edited by R. Busnel and J. F. Fish (Plenum, New York), pp. 291–297.
Zimmer, W. M. X., Johnson, M. P., D'Amico, A., and Tyack, P. L. (**2003**). "Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macrocephalus*)," IEEE J. Ocean. Eng. **28**, 13–28.
Zimmer, W. M. X., Johnson, M. P., Madsen, P. T., and Tyack, P. L. (**2005a**). "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)," J. Acoust. Soc. Am. **117**, 3919–3927.
Zimmer, W. M. X., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (**2005b**). "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," J. Acoust. Soc. Am. **117**, 1473–1485.

Automatic detection of marine mammals using information entropy

Christine Erbe[a)]
JASCO Applied Sciences Ltd., 55 Fiddlewood Crescent, Bellbowrie, Queensland 4070, Australia

Andrew R. King[b)]
CSIRO-QCAT, Technology Court, Pinjarra Hills, Queensland 4069, Australia

(Received 24 January 2008; revised 7 July 2008; accepted 14 August 2008)

This article describes an automatic detector for marine mammal vocalizations. Even though there has been previous research on optimizing automatic detectors for specific calls or specific species, the detection of any type of call by a diversity of marine mammal species still poses quite a challenge—and one that is faced more frequently as the scope of passive acoustic monitoring studies and the amount of data collected increase. Information (Shannon) entropy measures the amount of information in a signal. A detector based on spectral entropy surpassed two commonly used detectors based on peak-energy detection. Receiver operating characteristic curves were computed for performance comparison. The entropy detector performed considerably faster than real time. It can be used as a first step in an automatic signal analysis yielding potential signals. It should be followed by automatic classification, recognition, and identification algorithms to group and identify signals. Examples are shown from underwater recordings in the Western Canadian Arctic. Calls of a variety of cetacean and pinniped species were detected.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2982368]

PACS number(s): 43.30.Sf, 43.60.Bf [WWA] Pages: 2833–2840

I. INTRODUCTION

A. Background on automatic detection

Passive acoustics is becoming a frequently used tool in marine mammal surveys for population mapping, study of behavior, migration monitoring, and environmental impact assessments. With improving, more accessible, and cheaper ocean technology, the amount of data that can be collected and thus needs to be analyzed is increasing rapidly. Recording systems such as ocean bottom mounted hydrophones, which can be deployed for long time periods and which can record continuously for entire seasons, even years, produce vast amounts of data.[1–3] In July 2007, Urazghildiiev and Clark[4] summarized: "Currently, the detection of signals of interest is typically accomplished through a combination of visual inspection of spectrograms and listening to the data." This method is becoming impractical. However, techniques to automatically detect marine mammals are still not reliable and robust enough, particularly in situations of multiple and diverse target signals and diverse and varying background noise.

The performances of experienced human operators (EHOs) and an automatic detector that was based on the generalized likelihood ratio test (GLRT) have been compared for the task of detecting synthetic frequency-modulated signals mixed into Gaussian white noise.[4] Human operators simultaneously listened to the sounds and inspected spectrograms visually. The GLRT detector[5] outperformed the human operators. Using right whale calls in impulsive noise, this detector was later outperformed by an finite impulse response (FIR) filter operating on spectrograms.[6]

A matched filter, a hidden Markov model, and spectrogram cross-correlation have been used to detect bowhead whale calls in Arctic recordings, with the spectrogram cross-correlation yielding the best results.[7] The method was improved by measuring time variability within the call with a probability distribution function.[8] Matched filtering, spectrogram cross-correlation, and summed autocorrelation were subsequently optimized for other mysticete calls in various types of noise.[9,10] A feed-forward neural network outperformed the spectrogram cross-correlator for bowhead calls[11,12] and right whale calls.[13]

The performances of a matched filter, spectrogram cross-correlation, critical-band correlation, a back-propagation neural network, human listeners, and a trained whale were compared for the task of detecting beluga whale calls in ship noise, bubble noise, ice-cracking noise, and Gaussian noise.[14,15] The goal of these studies was to model masking in beluga whales. Therefore, performance was judged in terms of how closely the detection thresholds modeled those of the whale and not in terms of the lowest signal-to-noise ratio (SNR). Results differed for the different noises. On average, the neural net and the human detectors modeled the beluga thresholds best.

The performance of an automatic detector is influenced by the preprocessing of input data. When spectrographic features are extracted first, a matched filter working on feature vectors rather than pressure time series or spectrograms can be very successful.[16] Automatic detectors that work in real

[a)]Electronic mail: christine@jasco.com
[b)]Electronic mail: andrew.king@csiro.au

TABLE I. Call types of marine mammals likely encountered in the Canadian Arctic.

	Bandwidth (Hz)	Dominant frequency (Hz)	Duration (s)
Bowhead whales[a]			
FM tones	25–1200	100–400	0.4–3.8
Song	20–5000	<1000	minutes to hours
Pulsive	25–3500	<1000	0.3–7.2
Gray whales[b]			
Moans	20–2000	200–700	0.1–2
Pulsive	45–4500	300–800	0.01–0.2
FM tones	100–12 000	300	1–4
Clicks	100–10 000	1400–4000	0.002–0.05
Beluga whales[c]			
Whistles	300–18 000	1000–4000	0.01–3.8
Pulsed calls	200–20 000	<5000	0.03–4
Echolocation clicks	<12 0000	100 000–115 000	50–80 μs per click in a train
Narwhals[d]			
Whistles	300–18 000	500–6000	0.1–6
Pulsed calls	Narrow	Ranging from 500 to 24 000	0.5–2
Echolocation clicks	<160 000	40 000	<1 ms per click in a train
Walrus[e]			
Knocks	25–4000	100–1000	0.1–0.4
Bell calls	25–3000	500–750	0.8–2
Grunts	25–1000	100–500	0.2–0.7
Bearded seals[f]			
FM signals	200–11 000	1000–6000	1–86
Ringed seals[g]			
barks+yelps	100–6000	100–2000	0.2–1

[a]References 21–24.
[b]References 25–32.
[c]References 33–42.
[d]References 43–47.
[e]References 48–50.
[f]References 48, 51, and 52.
[g]Reference 48.

time in the field are sometimes based on measuring the peak energy above a moving average or the peak energy in a number of frequency bands.[17]

The information-theoretical concept of entropy is a measure of the amount of information contained in a signal. The entropy of acoustic power spectrum densities measures the peakedness or concentration of the energy distribution across frequency and has been used to characterize and identify animal signals.[18,19]

B. Motivation for the current study

JASCO Applied Sciences and one of the authors (C.E.) have undertaken acoustic measurements for environmental impact assessments of marine industrial activities in the Arctic for almost two decades. The scope of these projects is steadily increasing. In recent years, JASCO has been deploying dozens of autonomous hydrophone systems in the Arctic to monitor marine mammal population distributions and migrations before, during, and after industrial activities. Each year, terabytes of data need to be analyzed. Using EHOs exclusively has become impractical. An automatic tool for the detection and classification of marine mammal calls was needed.

While the majority of studies quoted above developed automatic detectors for one particular species, or, in fact, for only one specific call, the requirement of our tool was to detect as many marine mammals as possible across all species present in the area. This article investigates three tools for automatic detection of marine mammals and compares their performance.

C. Target species

Species found in the Western Canadian Arctic include bowhead whales (*Balaena mysticetus*), gray whales (*Eschrichtius robustus*), beluga whales (*Delphinapterus leucas*), narwhals (*Monodon monoceros*), polar bears (*Ursus maritimus*), walrus (*Odobenus rosmarus*), bearded seals (*Erignathus barbatus*), Largha (spotted) seals (*Phoca largha*), ringed seals (*Pusa hispida*), and possibly others.[20] Their call types are summarized in Table I. Searches of Cambridge Sci-

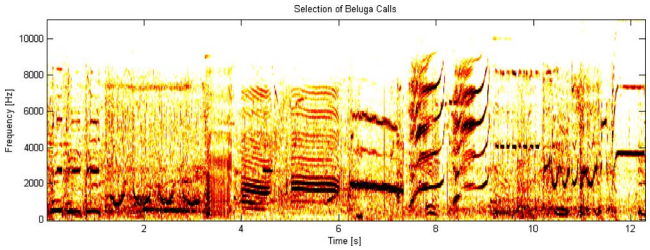


FIG. 1. (Color online) A series of beluga calls indicating the range in frequency and time characteristics.

entific Abstracts and ISI Web of Science (Science Citation Index) revealed no peer-reviewed publications of underwater vocalizations of polar bears and Largha seals.

Beluga whales are one of the most vociferous cetacean species, exhibiting a large vocal repertoire. Various attempts to classify their sounds have produced 20–50 distinct call types. A variety of beluga calls (recorded with a 10 kHz bandwidth) is shown in Fig. 1. The diverse frequency and time characteristics of beluga calls and the breadth of vocalizations of other Arctic animals posed a challenge for the automatic detector design.

II. METHODS

For the development and testing of the automatic detectors, 7 h of underwater recordings from the Western Canadian Arctic were used. These recordings were taken from a number of different studies by the authors and some of their colleagues over the past 20 years, reflecting a breadth of animals and ambient noise conditions. Ambient noise included wind, waves, shipping, distant seismics, cable noise, flow noise, ice noise, and bubble noise. Recording equipments differed. Sampling frequencies varied from 22 050 to 44 100 Hz. Some of the older recording equipment had a flat frequency response only up to about 9–10 kHz. All data were low-pass filtered in the frequency domain with a high-frequency cutoff at 8 kHz in order to unify the bandwidth across all recordings. Some of these studies had a marine mammal focus, while others had an entirely physical and acoustical oceanographic focus. Recordings in the latter case were not accompanied by visual observations of marine mammals. Therefore, it was generally not known which species were calling. This represents a realistic scenario where underwater recordings are obtained with autonomous systems deployed over extended periods of time. The current article deals only with signal detection and not with signal classification. All detectors were developed in MATLAB (© The MathWorks, Inc).

A. Peak energy detector

This method detected calls based on a simple energy criterion: sections of the time series with energy higher than the background by some threshold were regarded as being calls. The recorded pressure time series were high-pass filtered with a lower cutoff frequency of 20 Hz. Data were squared to yield an energy metric. The energy of the time series was averaged within short time windows, and each time window was tested for the presence of a call.

Calls were deemed to be present when the energy deviated sufficiently far from the background. To determine when this occurred, a characterization of the background noise energy in terms of an average and a variance from this average was needed. Under the assumption that animal calls were present for only a small portion of the recorded time series, noise statistics were computed without first having to remove the calls. Summed energy was computed over a series of time windows, and the median of the windowed energy was used as the noise level. The median rather than the mean was chosen as the energy average after initial trials and for the following reasons. In sections of the recorded time series that contained loud calls, the energy distribution curve had a long positive tail. Computing a mean places the energy average into the region of call energy, and quiet calls that did not surpass the mean by some threshold were missed. Taking the median energy on the other hand, places the energy average into the lower end of the energy distribution curve where mostly noise exists, thus allowing for the detection of quiet calls. The standard deviation of the energy time series was used to characterize the noise variance.

If the energy E in a time window was greater than the noise energy median $\bar{E}$ by more than a threshold γ times the standard deviation σ,

$$E > \bar{E} + \gamma \cdot \sigma,$$

then the window was deemed to contain a call. We used windows of 1 min length to compute the noise statistics. In order to reject sharp transient (cable) noise, which was present in some of the data sets, a signal had to be greater than the threshold for a minimum length of 50 ms in order to qualify as a call.

B. Multiband energy detector

The peak energy detector, depending on the threshold γ, either missed a lot of calls or gave a lot of false alarms. The majority of calls showed energy concentrated in a number of narrow frequency bands. These calls were audible to EHOs at very low levels because of their energy concentration. The peak energy detector was extended by first splitting the time series into a set of bandpassed versions at different frequencies. The peak energy algorithm was then separately run on each frequency band, and a signal was deemed present if the energy in any of the bands surpassed a threshold.

In detail, data were processed in 1 min blocks as above. The pressure time series were Fourier transformed and split into overlapping bands. Center frequencies were chosen in half-octave steps from 100 Hz to 8 kHz. Bands were one octave wide and thus overlapped. The shape of the filter used to split the signal into frequency bands was triangular in logarithmic frequency space. In analogy to the peak energy detector above, each of the bandpassed series was squared to yield an energy metric, and medians and standard deviations were computed. Using the same threshold criterion as above, if the instantaneous energy E surpassed the median by a threshold γ times the standard deviation σ in any one fre-

quency band, a call was deemed detected. Again, detections shorter than 50 ms were disregarded because of the occasional cable noise present.

C. Entropy detector

The entropy detector takes the multiband energy detector one step further. We were looking for a more generic measure of energy concentration. After examining a large number of recordings by ear and eye (from spectrograms), it became obvious that most animal calls across the various species had spectral characteristics that differed from background noise in the concentration or peakedness of spectral energy. Whistles and "moans" of the prevailing cetaceans and pinnipeds had clear fundamental tones with associated harmonics and sidebands. Other calls, such as "grunts," while more whitish and broadband, still had energy concentrated in certain frequency bands. Figure 1 illustrates the multipeaked nature of beluga call spectrograms. In contrast, typical noise sources, such as wind, waves, thermal icecracking, ship propeller noise, and seismic shots, had a much broader frequency distribution even if they were bandlimited as in the case of distant seismics with energy below a few hundred hertz.

Shannon entropy[53–56] measures the amount of information contained in a signal. It is used to characterize probability distributions in terms of their degree of peakedness. It is hence a candidate characteristic for marine mammal call detection. A probability distribution, which is strongly peaked, is one indicating a high degree of confidence in the value of the parameter being measured and is described as having low entropy. By contrast, a function that is very broad displays more uncertainty in the parameter value and is described as having high entropy.

If a parameter x can have one of N discrete values (or can lie in a set of discrete "bins") and our knowledge about its value is described by the set of probabilities p_k, being the probability that it has the kth value, then the entropy of this distribution is given by

$$H(p) = -\sum_{k=1}^{N} p_k \log p_k.$$

Note that the p_k's, being the probability distributions, must satisfy

$$\sum_{k=1}^{N} p_k = 1.$$

That is, the probability that the parameter has some value in its range is 1. In addition, probabilities are always positive, $0 \leqslant p_k \leqslant 1$. The entropy describes the degree to which the probabilities are concentrated onto a small number of the possible parameter values. Note that the parameter values do not appear in the entropy expression, so entropy does not distinguish between probability distributions with different numbers of peaks, provided that the degree of concentration of probabilities into bins is the same.

Acoustic power spectral densities (PSDs) are also positive additive distributions, and so they can be treated as if they were probability distributions. All that is required is that they be normalized by the total integrated power, so that they sum to 1. In detail, pressure time series were read in 1 min windows and high-pass filtered with a low-frequency cutoff at 20 Hz. The mean PSD was computed over the entire 1 min window and used to prewhiten the data. In a series of shorter, 60 ms long, windows, PSDs were computed and normalized by the mean PSD. Entropy was computed as the product of the PSD and the logarithm of the PSD. The normalization makes the entropy independent of the total energy; i.e., a quiet call has the same entropy as a loud call of the same type. This yielded a time series of entropies, each computed in 60 ms long windows. The median and variance were computed over the longer 1 min window. Wherever the instantaneous entropy surpassed the median by a threshold γ times the standard deviation σ, a call was deemed detected.

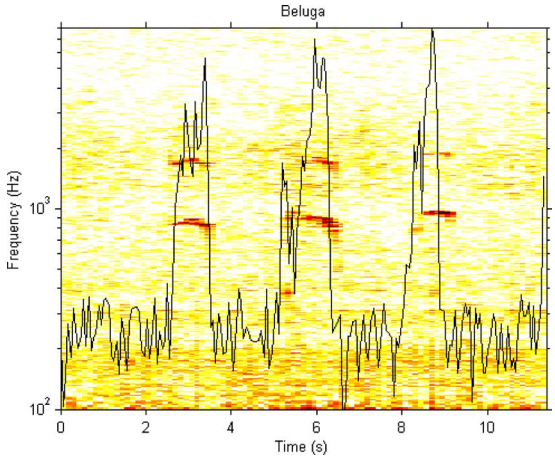


FIG. 2. (Color online) Spectrogram of a beluga call with entropy curve overlain. The entropy (not calibrated hence scales less) is high at the times of the call pulses and low in between. A detection threshold as a horizontal line in the picture should ideally lie above the background entropy and below the maximum call entropy.

D. Experienced human operators

A human is considered an EHO after lengthy periods of spectrogram inspection and listening.[4] EHOs were used to decide which detections were correct detections or false alarms and to find missed detections. All 7 h of recordings were analyzed by listening and visual discrimination of spectrograms. The majority of calls were very faint, and EHOs were able to "rewind" in order to listen to potential signals multiple times before making a decision. A total of 213 calls were detected from various species. No attempt was made to classify them or identify species. We looked for marine mammal vocalizations in general.

The three automatic detectors described above were then run on the 7 h of recordings at a low threshold γ, deliberately yielding a high number of correct detections at the expense of a high number of false alarms. All false alarms of the three detectors were subsequently investigated by the same EHOs by ear and eye. A number of very quiet calls and, in particular, short calls were thus detected, which had

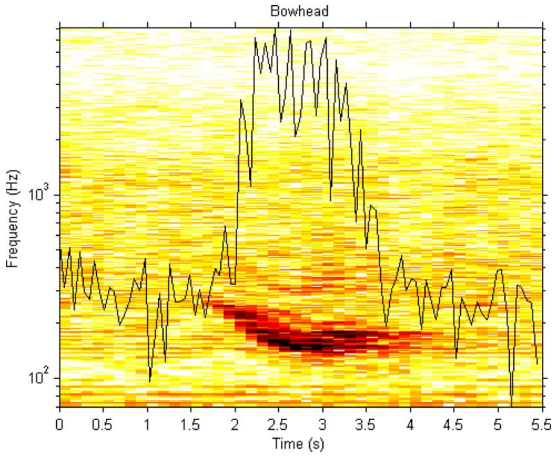


FIG. 3. (Color online) Spectrogram of a bowhead call with entropy curve overlain.

previously been missed by the same EHOs. The total number of calls detected by the EHOs with prompting by the automatic detectors rose to 297.

III. RESULTS

A. Examples from the entropy detector

All of the three detectors computed a time series of a statistic of the data and defined detections at locations where this function surpassed a preset threshold. A good statistic will show clear differences between background noise and signal with noise present, so that it is relatively insensitive to the threshold value chosen.

Results are plotted only for the entropy detector. Figure 2 shows the spectrogram of a faint beluga whale call and the associated entropy curve. Entropy was not calibrated and is thus given without scales. A good detector would have a threshold above the background entropy and below the call entropy. Figure 3 shows the spectrogram of a bowhead call with its associated entropy, similar to a published "glide."[23] Figure 4 sounded like a walrus grunt to the EHOs, and the time and frequency characteristics matched published spectrograms of walrus.[48]

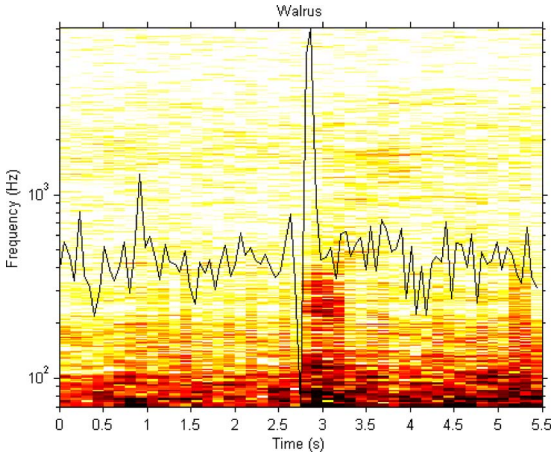


FIG. 4. (Color online) Probably a walrus grunt and associated entropy.

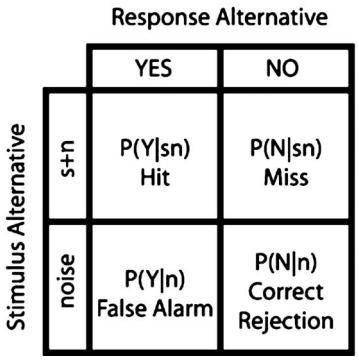


FIG. 5. Contingency table for the automatic detection problem.

B. ROC curves

Signal detection theory was used to compare the performances of the three automatic detectors.[57,58] The automatic detection tasks reported here are mathematically equivalent to a "two-class prediction" or "binary classification" problem. There are four possible outcomes (Fig. 5). With $P(Y|\mathrm{sn})$ and $P(N|\mathrm{sn})$ denoting the probabilities of responding "yes" or "no" to a signal in noise, the following relationships exist, leaving only two of the four probabilities in Fig. 5 independent: $P(Y|\mathrm{sn})+P(N|\mathrm{sn})=1$ and $P(Y|\mathrm{n})+P(N|\mathrm{n})=1$.

In receiver-operating-characteristic (ROC) plots, the probability of false alarms is plotted versus the probability of correct detections. If one changes the threshold γ of a detector, so-called ROC curves are produced. This was done in Fig. 6. The "perfect" detector would have a hit rate of 1 and a false alarm rate of 0 and would thus be a point on the ROC plot in the top left corner at (0,1). The detector that approaches this point most closely is the "best" detector, in this case the entropy detector.

The probability of correct detection was computed by comparing the time stamps of the automatic detections with those of the EHOs. For each of the EHO detections, a detection window was defined as the duration of the call plus 1 s before and after. If the automatic detector showed any number of detections within this window, this was considered as one correct detection. The number of correct detections was divided by the total number of EHO detections to yield the

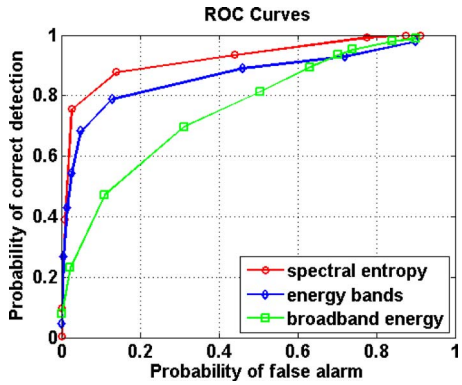


FIG. 6. (Color online) Varying the threshold γ for each detector yielded ROC curves. The total number of signals was 297. The best detector is the one reaching closest to the point (0,1); i.e., here the best is the entropy detector.

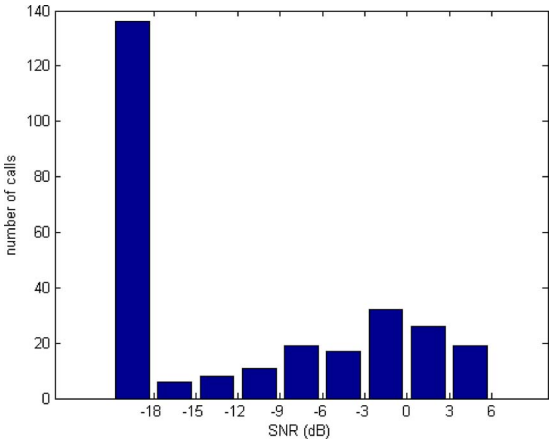


FIG. 7. (Color online) Distribution of SNRs for all 297 signals.

probability of correct detection for each automatic tool. All automatic detections that fell outside of the EHO detection windows were considered false alarms. The maximum possible number of false alarms is the number of samples in the time series of the detection metric (energy or entropy) minus the total number of samples in all EHO detection windows. This leaves the number of samples in the time series outside of the detection windows, which could potentially be labeled as a false alarm. The probability of false alarms is then the total number of false alarms divided by the maximum possible number of false alarms.

C. Signal-to-noise ratios

The SNR indicates how difficult the detection task was. Almost all of the signals had energy spread over a number of differing frequency bands (harmonics, overtones, and sidebands). For all of the signals, the main energy lay between 20 Hz and 3 kHz. This band was used to compute SNRs. The mean power of signal+noise was computed over the duration of the signal. The mean noise power was computed from a 1 s window prior to the signal onset. Subtracting the latter from the former yields an estimate of the signal power. The SNR is the ratio of signal-to-noise power, expressed in decibel. Figure 7 shows the number of calls per SNR. The majority of calls were very faint. However, one has to bear in mind that SNR was computed over a 3 kHz band, and signal and noise did not occupy the exact same frequencies in most cases, aiding detection in particular at the negative SNRs. As the noise level varied with time, SNRs less than about −15 dB could not be estimated reliably.

Table II lists the number of hits and false alarms as a function of threshold and SNR for the entropy detector. The number of true signals depends on the SNR but not on the threshold. The number of false alarms depends on the threshold but not on the SNR. The hit rate is the ratio of the number of hits and the number of true signals. At a low threshold (0.5), the hit rate was high: 168 of the 187 faint signals with a SNR of <−5 dB were detected (hit rate =168/187), and all of the 26 loud signals with a SNR of >5 dB were detected (hit rate=1). The hit rate was higher, the higher the SNR. The number of false alarms at this low threshold was high, too: 3627. The number of hits decreased with increasing threshold, as did the number of false alarms.

TABLE II. Number of hits and false alarms for the entropy detector as a function of threshold and SNR.

		Threshold γ					No. of true signals
		0.5	1	2	4	8	
SNR (dB)	<−5	168	157	134	64	13	187
	−5–0	39	35	32	17	6	42
	0–5	36	35	29	16	6	37
	>5	26	25	22	12	4	26
No. of false alarms		3627	538	0	0	0	

IV. DISCUSSION

This article presented a method for automatic detection of marine mammal sounds in large data sets. Unlike previous studies, which focused on detecting one particular species, sometimes even just one particular vocalization, the goal of this study was to detect all vocalizing marine mammals at even the quietest levels. After an initial investigation by EHOs who listened to recordings while observing spectrograms in real time, it was found that most noises had broad (even if limited to low frequencies) or unstructured spectra, while most calls consisted of multiple peaks (harmonics, overtones, and sidebands). Entropy is sensitive to this difference and was therefore used to design an automatic detector. Performance was compared to that of a simple peak-energy detector and a multiband energy detector. The performance of all three detectors was judged by EHOs who investigated all false alarms and searched for missed calls by listening and through visual spectrogram discrimination. The entropy detector outperformed the other two, as evidenced by their ROC curves.

Ambient noise is typically a nonstationary random process with unknown statistical and spectral properties. While the methods based on detecting peak energy relied on computing statistical moments around a mean, entropy does not make any underlying statistical assumptions—except that noise has a "less structured" spectrum than signal. False alarms for all three detectors were due to shipping, distant seismics, cable noise, flow noise, and bubble noise. The entropy detector excels at finding tonal signals or signals where the energy is concentrated in a limited number of frequency bands. Shipping noise, although broadband, can exhibit strong tonal peaks below 1 kHz. The entropy detector was made robust to shipping noise by prewhitening the data in 1 min long sliding windows. The type of noise that caused the most problems for the entropy detector was bubble noise. There were clear tonal "blub" sounds in some of the recordings. These were likely bioacoustic in origin but were considered noise.

Some of the animals encountered in the Arctic produce clicks (Table I). However, the EHOs did not identify any

click signals in the recordings. It is expected that the entropy detector would perform poorly on short and broadband clicks.

The entropy detector and the multiband energy detector—at an appropriately low threshold—found about 30% more quiet and short calls than did the EHOs at first pass in real time. This highlights the problems with attitude, bias, and concentration when using EHOs. ROC curves can be computed for humans and animals,[59] allowing an assessment of attitude and bias,[57] but this was not done here. One also needs to decide how useful and necessary the detection of extremely faint signals is. The answer depends on the goal of the project. In areas that might constitute a "critical" habitat for marine mammals and where marine mammals are (logistically) difficult to detect and keep track of with methods other than passive acoustics, it might well be necessary to detect even the furthest and quietest calls.

The above methods could be improved by designing an adaptive algorithm to compute noise statistics. The methods all relied on the computation of a median of a statistical metric within a sliding time window. For the two energy detectors, if the instantaneous energy surpassed the median energy by a threshold times the standard deviation, a call was deemed present. In the case of the entropy detector, the mean power spectrum density over a predefined averaging window was used to prewhiten the data, and instantaneous entropy was compared to the median entropy. The averaging time window has to be short enough for the background statistics to adjust to temporary changes in noise energy. For example, if a boat passes by and the time window is long, the boat will be detected as a brief increase in energy and perhaps entropy and will thus give a false alarm. If the time window is short, the energy of the boat will fall into the background noise statistic, and quiet calls above the boat energy will be detected correctly. The opposite problem occurs when a number of animals vocalize simultaneously for some time. If the averaging time window is short, the call energy will become part of the background, yielding a missed detection. The time window would have to be long enough to pick the "conversation" up as a correct detection. The implementation of an adaptive window would help in these situations. However, it might be computationally more efficient to use the fast methods described in this article as a first pass and to sort out false alarms by a more specific classification and recognition algorithm later. Another idea is to accumulate noise statistics only during times when a signal is not detected, so when multiple animals are calling, noise characterization is suspended until the animals are quiet. This would require simultaneous detection and classification.

There are various ways in which the above detectors can be improved. However, further improvements depend on the specific goals and limitations of the study, as well as the types of signals and noises present. If a large variety of signals is to be detected and if computational speed is important, then the above methods can yield potential detections considerably faster than real time. A 32 min file (.wav) was processed in just under 1 min on a four-year-old laptop. The automatic detectors presented in this article were intended as a first step in automatic signal analysis. As a second step, all detected signals should be passed to an automatic classification algorithm that groups them. Automatic recognition or identification algorithms could then search for particular species or particular types of noise. Tasks such as "identifying all animals that were present prior to, during and after industrial activity," or "locating and tracking bowheads during their migration past industrial sites" could then be handled.

V. CONCLUSION

With improving and cheaper technology and with increasing scope of acoustic monitoring studies, the amount of data collected rises, rendering EHOs as detectors and classifiers impractical. There is an increasing need for efficient, reliable, and fast automatic tools. The entropy detector presented here detected calls from a variety of marine mammal species in diverse ambient noise conditions and outperformed two types of energy detectors. It is used as a first step in automatic signal analysis and should be followed by classification, recognition, and identification algorithms.

ACKNOWLEDGMENTS

Underwater sounds in the Arctic were recorded by the Freshwater Institute, Winnipeg, the Institute of Ocean Sciences, Sidney BC, the Vancouver Aquarium, and the University of British Columbia. In particular, we thank Sue Cosens, John Ford, David Farmer, and Yunbo Xie. We acknowledge the two very careful and constructive reviewers of this article.

[1]C. G. Fox, H. Matsumoto, and T. K. A. Lau, "Monitoring Pacific Ocean seismicity from an autonomous hydrophone array," J. Geophys. Res. **106**, 183–4206 (2001).
[2]C. W. Clark, F. Borsani, and G. Notarbartolo-di-Sciara, "Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea," Marine Mammal Sci. **18**, 286–295 (2002).
[3]M. O. Lammers, R. E. Brainard, W. W. L. Au, T. A. Mooney, and K. B. Wong, "An ecological acoustic recorder (EAR) for long-term monitoring of biological and anthropogenic sounds on coral reefs and other marine habitats," J. Acoust. Soc. Am. **123**, 1720–1728 (2008).
[4]I. R. Urazghildiiev and C. W. Clark, "Detection performances of experienced human operators compared to a likelihood ratio based detector," J. Acoust. Soc. Am. **122**, 200–204 (2007).
[5]I. Urazghildiiev and C. Clark, "Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test," J. Acoust. Soc. Am. **120**, 1956–1963 (2006).
[6]I. R. Urazghildiiev and C. W. Clark, "Acoustic detection of North Atlantic right whale contact calls using spectrogram-based statistics," J. Acoust. Soc. Am. **122**, 769–776 (2007).
[7]D. K. Mellinger and C. W. Clark, "A method for filtering bioacoustic transients by spectrogram image convolution," Proc. IEEE **3**, 122–127 (1993).
[8]D. K. Mellinger, "Handling time variability in bioacoustic transient detection," Oceans '93, Engineering in Harmony with the Ocean, Proc. IEEE **3**, 116–121 (1993).
[9]D. K. Mellinger and C. W. Clark, "Methods for automatic detection of mysticete sounds," Mar. Freshwater Behav. Physiol. **29**(3), 163–181 (1997).
[10]L. M. Munger, D. K. Mellinger, S. M. Wiggins, S. E. Moore, and J. A. Hildebrand, "Performance of spectrogram cross-correlation in detecting right whale calls in long-term recordings from the Bering Sea," Can. Acoust. **33**, 25–34 (2005).
[11]J. R. Potter, D. K. Mellinger, and C. W. Clark, "Marine mammal call discrimination using artificial neural networks," J. Acoust. Soc. Am. **96**, 1255–1262 (1994).
[12]D. K. Mellinger and C. W. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**,

3518–3529 (2000).
[13]D. K. Mellinger, "A comparison of methods for detecting right whale calls," Can. Acoust. **32**, 55–65 (2004).
[14]C. Erbe, A. R. King, M. Yedlin, and D. M. Farmer, "Computer models for masked hearing experiments with beluga whales (*Delphinapterus leucas*)," J. Acoust. Soc. Am. **105**, 2967–2978 (1999).
[15]C. Erbe, "Detection of whale calls in noise: Performance comparison between a beluga whale, human listeners, and a neural network," J. Acoust. Soc. Am. **108**, 297–303 (2000).
[16]K. M. Stafford, C. G. Fox, and D. S. Clark, "Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean," J. Acoust. Soc. Am. **104**, 3616–3625 (1998).
[17]D. Gillespie and O. Chappell, "An automatic system for detecting and classifying the vocalizations of harbour porpoises," Bioacoustics **13**, 37–61 (2002).
[18]D. Valente, H. Wang, P. Andrews, P. P. Mitra, S. Saar, O. Tchernichovski, I. Golani, and Y. Benjamini, "Characterizing animal behavior through audio and video signal processing," IEEE Multimedia **14**, 32–41 (2007).
[19]D. K. Mellinger and J. W. Bradbury, "Acoustic measurement of marine mammal sounds in noisy environments," Proceedings of the Second International Conference on Underwater Acoustic Measurements: Technologies and Results, Heraklion, Crete (25–29 June 2007), 8 pp.
[20]R. R. Reeves, B. S. Stewart, P. J. Clapham, and J. A. Powell, *National Audubon Society Guide to Marine Mammals of the World* (Chanticleer, 2002).
[21]C. W. Clark and J. H. Johnson, "The sounds of the bowhead whale, *Balaena mysticetus*, during the spring migrations of 1979 and 1980," Can. J. Zool. **62**, 1436–1441 (1984).
[22]B. Wursig and C. Clark, "Behavior," in *The Bowhead Whale*, edited by J. J. Burns, J. J. Montague, and C. J. Cowles (Society for Marine Mammalogy, 1993), pp. 157–199.
[23]D. K. Ljungblad, P. O. Thompson, and S. E. Moore, "Underwater sounds recorded from migrating bowhead whales, *Balaena mysticetus*, in 1979," J. Acoust. Soc. Am. **71**, 477–482 (1982).
[24]W. C. Cummings and D. V. Holliday, "Sounds and source levels from bowhead whales off Pt. Barrow, AK," J. Acoust. Soc. Am. **82**, 814–821 (1987).
[25]N. L. Crane and K. Lashkari, "Sound production of gray whales, *Eschrichtius robustus*, along their migration route: A new approach to signal analysis," J. Acoust. Soc. Am. **100**, 1878—1886 (1996).
[26]F. J. Ollervides, "Gray whales and boat traffic: Movement, vocal, and behavioral responses in Bahia Magdalena, Mexico," Ph.D. thesis, Texas A&M University (2001).
[27]K. M. Stafford, S. E. Moore, M. Spillane, and S. Wiggins, "Gray whale calls recorded near Barrow, Alaska, throughout the winter of 2003-04," Arctic **60**, 167–172 (2007).
[28]S. E. Moore, K. M. Stafford, D. K. Mellinger, and J. A. Hildebrand, "Listening for large whales in the offshore waters of Alaska," BioScience **56**, 49–55 (2006).
[29]S. E. Moore and D. K. Ljungblad, "Gray whales in the Beaufort, Chukchi and Bering Seas: Distribution and sound production," in *The Gray Whale, Eschrichtius robustus*, edited by M. L. Jones, S. L. Swartz, and S. Leatherwood (Academic, New York, 1984), pp. 543–559.
[30]W. C. Cummings, P. O. Thompson, and R. Cook, "Underwater sounds of migrating gray whales, *Eschrichtius glaucus* (Cope)," J. Acoust. Soc. Am. **44**, 1278–1281 (1968).
[31]J. F. Fish, J. L. Sumich, and G. L. Lingle, "Sounds produced by the gray whale, *Eschrichtius robustus*," Mar. Fish. Rev. **36**, 38–45 (1974).
[32]M. E. Dahlheim, "Bio-acoustics of the gray whale (*Eschrichtius robustus*)," M.S. thesis, University of British Columbia (1987).
[33]R. A. Belikov and V. M. Bel'kovich, "Underwater vocalization of the white whales (*Delphinapterus leucas*) in a reproductive gathering during different behavioral situations," Okeanologiya **43**, 118–126 (2003).
[34]J. D. Karlsen, A. Bisther, D. Ludersen, T. Haug, and K. M. Kovacs, "Summer vocalisations of adult male white whales (*Delphinapterus leucas*) in Svalbard, Norway," Polar Biology **25**, 808–817 (2002).
[35]B. L. Sjare and T. G. Smith, "The vocal repertoire of white whales, *Delphinapterus leucas*, summering in Cunningham Inlet, Northwest Territories," Can. J. Zool. **64**, 407–415 (1986).
[36]S. M. VanParijs, C. Lydersen, and K. M. Kovacs, "Sounds produced by individual white whales, *Delphinapterus leucas*, from Svalbard during capture," J. Acoust. Soc. Am. **113**, 57–60 (2003).
[37]A. Faucher, "The vocal repertoire of the St. Lawrence Estuary population of beluga whale (*Delphinapterus leucas*) and its behavioral, social and environmental contexts," M.S. thesis, Dalhousie University, Canada (1988).
[38]N. M. Angiel, "The vocal repertoire of the beluga whale in Bristol Bay, Alaska," M.S. thesis, University of Washington, USA (1997).
[39]W. E. Schevill and B. Lawrence, "Underwater listening to the white porpoise (*Delphinapterus leucas*)," Science **109**, 143–144 (1949).
[40]M. P. Fish and W. H. Mowbray, "Production of underwater sound by the white whale or beluga, *Delphinapterus leucas*," J. Mar. Res. **20**, 149–162 (1962).
[41]W. W. L. Au, D. A. Carder, R. H. Penner, and B. L. Scronce, "Demonstration of adaptation in beluga whale echolocation signals," J. Acoust. Soc. Am. **77**, 726–730 (1985).
[42]W. W. L. Au, R. H. Penner, and C. W. Turl, "Propagation of beluga echolocation signals," J. Acoust. Soc. Am. **82**, 807–813 (1987).
[43]J. K. B. Ford and H. D. Fisher, "Underwater acoustic signals of the narwhal *Monodon monoceros*," Can. J. Zool. **56**, 552–560 (1978).
[44]W. A. Watkins, W. E. Schevill, and C. Ray, "Underwater sounds of *Monodon* (Narwhal)," J. Acoust. Soc. Am. **49**, 595–599 (1971).
[45]B. Møhl, A. Surlykke, and L. A. Miller, "High intensity narwhal clicks," in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York, 1990), pp. 295–303.
[46]L. A. Miller, J. Pristed, B. Møhl, and A. Surlykke, "The clicksounds of narwhals (*Monodon monoceros*) in Inglefield Bay, Northwest Greenland," Marine Mammal Sci. **11**, 491–502 (1995).
[47]A. D. Shapiro, "Preliminary evidence for signature vocalizations among free-ranging narwhals (*Monodon monoceros*)," J. Acoust. Soc. Am. **120**, 1695–1705 (2006).
[48]I. Stirling, W. Calvert, and H. Cleator, "Underwater vocalizations as a tool for studying the distribution and relative abundance of wintering pinnipeds in the high Arctic," Arctic **36**, 262–274 (1983).
[49]B. Sjare, I. Stirling, and C. Spencer, "Structural variation in the songs of Atlantic walruses breeding in the Canadian High Arctic," Aquat. Mamm. **29**, 297–318 (2003).
[50]I. Stirling, W. Calvert, and C. Spencer, "Evidence of stereotyped underwater vocalizations of male Atlantic walruses (*Odobenus rosmarus rosmarus*)," Can. J. Zool. **65**, 2311–2321 (1987).
[51]H. J. Cleator and I. Stirling, "Winter distribution of bearded seals (*Erignathus barbatus*) in the Penny Strait area, Northwest Territories, as determined by underwater vocalizations," Can. J. Fish. Aquat. Sci. **47**, 1071–1076 (1990).
[52]D. Risch, C. W. Clark, P. J. Corkeron, A. Elephandt, K. M. Kovacs, C. Lydersen, I. Stirling, and S. M. VanParijs, "Vocalizations of male bearded seals, *Erignathus barbatus*: classification and geographical variation," Anim. Behav. **73**, 747–762 (2007).
[53]E. T. Jaynes, "Information theory and statistical mechanics," Phys. Rev. **106**, 620–630 (1957).
[54]D. S. Sivia, *Data Analysis: A Bayesian Tutorial* (Clarendon, Oxford, 1996).
[55]C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423 (1948); **27**, 623–656 (1948).
[56]C. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Champaign, IL, 1963).
[57]D. M. Green and J. M. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
[58]K. A. Spackman, "Signal detection theory: Valuable tools for evaluating inductive learning," Proceedings of the of Sixth International Workshop on Machine Learning, San Mateo, CA (1989), pp. 160–163.
[59]C. Erbe and D. M. Farmer, "Masked hearing thresholds of a beluga whale (*Delphinapterus leucas*) in icebreaker noise," Deep-Sea Res., Part II **45**, 1373–1388 (1998).

Waveguide invariant broadband target detection and reverberation estimation

Ryan Goldhahn,[a] Granger Hickman,[b] and Jeffrey Krolik[c]
Duke University, Durham, North Carolina 27708-0291

(Received 10 January 2008; revised 31 July 2008; accepted 1 August 2008)

Reverberation often limits the performance of active sonar systems. In particular, backscatter off of a rough ocean floor can obscure target returns and/or large bottom scatterers can be easily confused with water column targets of interest. Conventional active sonar detection involves constant false alarm rate (CFAR) normalization of the reverberation return which does not account for the frequency-selective fading caused by multipath propagation. This paper presents an alternative to conventional reverberation estimation motivated by striations observed in time-frequency analysis of active sonar data. A mathematical model for these reverberation striations is derived using waveguide invariant theory. This model is then used to motivate waveguide invariant reverberation estimation which involves averaging the time-frequency spectrum along these striations. An evaluation of this reverberation estimate using real Mediterranean data is given and its use in a generalized likelihood ratio test based CFAR detector is demonstrated. CFAR detection using waveguide invariant reverberation estimates is shown to outperform conventional cell-averaged and frequency-invariant CFAR detection methods in shallow water environments producing strong reverberation returns which exhibit the described striations.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2980519]

PACS number(s): 43.30.Vh, 43.30.Gv, 43.60.Uv [RCG] Pages: 2841–2851

I. INTRODUCTION

Estimating and normalizing reverberation is of considerable importance in active sonar. Shallow-water environments present significant challenges due to multipath propagation. Unlike deep water channels, nearly all propagation paths in shallow water are bottom-reflecting. Thus, a large portion of the signal energy interacts with the seafloor, which can produce a strong reverberation return. This motivates the development of active sonar signal processing for detection of targets that are designed to handle multipath-prone and reverberation-limited shallow water environments. Conventional active sonar detection involves basebanding, matching filtering, and normalizing the received time series. The envelope is then compared against a threshold. Cell-average constant false alarm rate (CA-CFAR) normalization, derived without consideration of multipath propagation, is traditionally used to estimate the threshold above which targets are discriminated against reverberation and noise. The threshold is estimated from neighboring data in the time series, which is assumed to contain only reverberation and noise with the same statistics as the data under test. By not accounting for propagation effects, the implicit assumptions made in CA-CFAR detection are often incorrect. In particular, narrowband returns, while exhibiting a small reverberation-induced variance, often exhibit a high degree of nonstationarity due to coherent constructive and destructive interference of propagation modes. On the other hand, wideband returns have more incoherent reverberation components which result in higher background variance without a concomitant increase in the target peak due to multipath delay spread.[1] Exploiting larger bandwidths in active sonar has thus involved a trade off between detection in nonstationary subbands versus broadband detection in reverberation with larger variance.

Reverberation modeling, with the ultimate goal of improving active sonar detection and classification, has been addressed via both statistical and physics-based approaches in the literature. The statistical approach seeks to go beyond the traditional assumption of reverberation statistics, i.e., that the return at any time sample is due to backscatter from a sufficient number of scatterers so as to make the sum complex Gaussian distributed (with a Rayleigh envelope) by virtue of the central limit theorem. Non-Rayleigh reverberation is often observed in real data, and work has been done to quantify the "non-Rayleighness."[2,3] The *K*-distribution has been found to be a useful statistical model for active sonar reverberation,[4,5] with considerable effort having been made to relate environmental and system parameters to the parameters of this distribution.[6,7] Physics-based reverberation modeling exploits acoustic scattering and propagation models and uses these along with knowledge of the environment and transmitted signal to predict reverberation levels. Acoustic propagation has historically been modeled using the ray approximation,[8] wavenumber integration,[9] and using the spectral integral,[10] however, normal mode methods are most commonly employed in shallow water using low frequencies.[11–13] This work models the propagation of the pulse to and from the scattering element using normal modes and employs an empirical Lambertian scattering model. More sophisticated scattering models have been proposed,[14–16] however, these require a characterization of the scattering surface. All methods of propagation and scat-

[a]Electronic mail: rag15@ee.duke.edu
[b]Electronic mail: gwh@ee.duke.edu
[c]Electronic mail: jk@ee.duke.edu

tering modeling are limited by the knowledge of the environment, and parameters of the waveguide and scattering surface are generally not known in sufficient detail to allow accurate long range reverberation prediction from modeling alone.

This work adopts a physics-based approach to reverberation modeling, however, reverberation estimation and target detection is based on the short-time Fourier transform of the data. Striations in time (or slant range) and frequency have been observed in the spectrograms of broadband acoustic propagation. The striations are regions of constant intensity, parameterized by a scalar known as the waveguide invariant, β.[17–19] The waveguide invariant is approximately constant in many general environments[20] and is approximately unity in shallow water waveguides. The value may vary spatially if the waveguide is not range-independent.[18] Recently, these striations have been used in ping-to-ping target tracking in bistatic active sonar.[21,22] Striations have also been observed in real[23] and simulated[24] reverberation data. The expected intensity is shown to be constant along curves parameterized by β. In this paper, a frequency-dependent reverberation estimate is derived that accounts for the striations induced by the channel. The reverberation level is estimated by averaging along the striations, rather than data adjacent in range. This yields an estimate with lower mean square error (MSE) and improved detection performance over conventional CA-CFAR techniques. In this paper, a normal mode based waveguide invariant reverberation model is derived, followed by a discussion of its application to reverberation estimation. The mean square error of the resulting waveguide invariant reverberation estimate is computed analytically. This estimate is then used in a generalized likelihood ratio test (GLRT) to obtain a waveguide invariant constant false alarm rate (WI-CFAR) detector. Results are presented for waveguide invariant reverberation estimation and CFAR detection using both real and simulated Mediterranean active sonar data.

II. REVERBERATION MODELING IN A SHALLOW-WATER WAVEGUIDE

This section describes how a statistical model for bottom backscatter can be combined with normal mode propagation modeling to characterize reverberation in a shallow-water waveguide. Below several kilohertz in shallow water, acoustic propagation can be modeled using a normal mode sum[1,11–13] which expresses the pressure field $p(r,z)$ at range and depth (r,z) due to a point source at the origin and depth z_s in terms of depth-dependent eigenfunctions $Z_m(z)$ and horizontal wavenumbers k_m,

$$p(\omega,r,z)=\sum_{m=1}^{\infty} Z_m(z_s)Z_m(z)\frac{\exp(-jk_m r)}{\sqrt{k_m r}}. \tag{1}$$

For active sonar a model for the return from a single scattering element which includes the propagation to the scatterer and back to the receiver must be considered. Assuming a monostatic geometry (i.e., the transmitter and receiver are colocated), the equation above becomes a double sum in the two-way frequency-domain Green's function

$$\begin{aligned} p(\omega,r,z) &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} Z_m(z_s)Z_m(z)S_{mn}Z_n(z)Z_n(z_s) \\ &\quad\times\frac{\exp(-j(k_m+k_n)r)}{r\sqrt{k_m k_n}} \\ &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} A_{mn}\frac{\exp(-j(k_m+k_n)r)}{r\sqrt{k_m k_n}}. \end{aligned} \tag{2}$$

The complex scalars S_{mn} are elements of the scattering matrix $\mathbf{S}$, which describes the coupling of modes by the scatterer, i.e., the element $S_{mn}=[\mathbf{S}]_{mn}$ specifies the amount of energy incident on mode m that is scattered onto mode n. The scattering and depth-dependent terms have been grouped into modal amplitude coefficients A_{mn}, which are assumed to be frequency-independent over the bandwidth of the acoustic source.

If (r,z) represents a point scatterer on the ocean floor, the frequency-domain received signal can be modeled by the Fourier component of the source signal, $U(\omega)$, scaled by a complex amplitude $\eta(r)$ and the two-way Green's function of Eq. (2),

$$p(\omega,r)=U(\omega)\sum_{m,n}\eta(r)A_{mn}\frac{\exp(-j(k_m+k_n)r)}{r\sqrt{k_m k_n}}. \tag{3}$$

Since only the return from the ocean floor is considered here, the depth dependence of the scatterer in Eq. (3) is suppressed for notational convenience. The scattering matrix $\mathbf{S}$ included in A_{mn} is generated according to a Lambertian scattering model where the (m,n)th element of $\mathbf{S}$ is proportional to the arrival angle of modes m and n, i.e., $S_{mn}=\sqrt{\sin(\phi_m)\sin(\phi_n)}$. Assuming $\eta(r)$ is frequency-independent, the corresponding time series due to a single bottom scatterer can be found from Eq. (3) by Fourier synthesis

$$\begin{aligned} p(t,r) &= \eta(r)\int_{-\infty}^{\infty} U(\omega)\left[\sum_{m,n} A_{mn}\frac{\exp(-j(k_m+k_n)r)}{r\sqrt{k_m k_n}}\right] \\ &\quad\times\exp(j\omega t)\mathrm{d}\omega. \end{aligned} \tag{4}$$

This is the time series return from a single scatterer. A complete reverberation data model can now be constructed by considering the returns from a dense grid of randomly distributed point scatterers. In this paper, the scattering amplitude $\eta(r)$ as a function of range is assumed to be a zero-mean complex Gaussian white random process such that $E[\eta(r)\eta^*(r')]=\sigma_\eta^2\delta(r-r')$. The reverberation time series is thus obtained by integrating Eq. (4) over range to obtain

$$\begin{aligned} y(t) &= \int_0^{\infty}\eta(r)\left[\int_{-\infty}^{\infty} U(\omega)\left[\sum_{m,n} A_{mn}\frac{\exp(-j(k_m+k_n)r)}{r\sqrt{k_m k_n}}\right]\right. \\ &\quad\left.\times\exp(j\omega t)\mathrm{d}\omega\right]\mathrm{d}r. \end{aligned} \tag{5}$$

Figure 1 shows an example of a simulated time-domain reverberation return based on the model in Eq. (5) for a propagation environment which will be discussed in more detail in Sec. V. Of particular interest here is the short-time

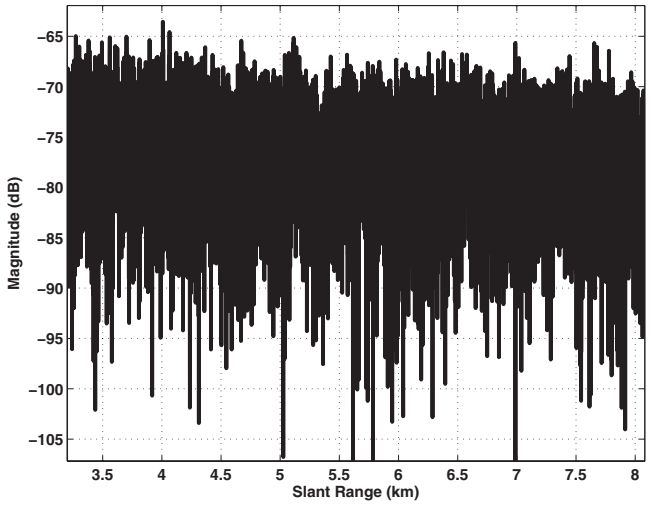


FIG. 1. Simulated time-domain reverberation return from a 400–1400 Hz signal. A weak target is present at 5.1 km, which is indistinguishable from reverberation-induced peaks at other ranges.

Fourier transform (STFT) of the time series return which is taken using a windowing function, $w(t)$, of width T and centered at time τ,

$$y(\omega,\tau,T)=\int_0^\infty\int_{-\infty}^\infty w\left(\frac{t-\tau}{T}\right)p(t,r)\exp(-j\omega t)\mathrm{d}t\mathrm{d}r. \quad (6)$$

A realization of the STFT magnitude is shown in Fig. 2, in which the time axis is plotted in terms of slant range under the formula $r=(ct)/2$, where c is the average sound speed in the water column. Observe that in this representation of the reverberation return, time-frequency striations of approximately constant average magnitude can be seen. This effect has been observed in both real and simulated data.[23,24] This phenomenon can be explained by waveguide invariant theory, described in the following section.

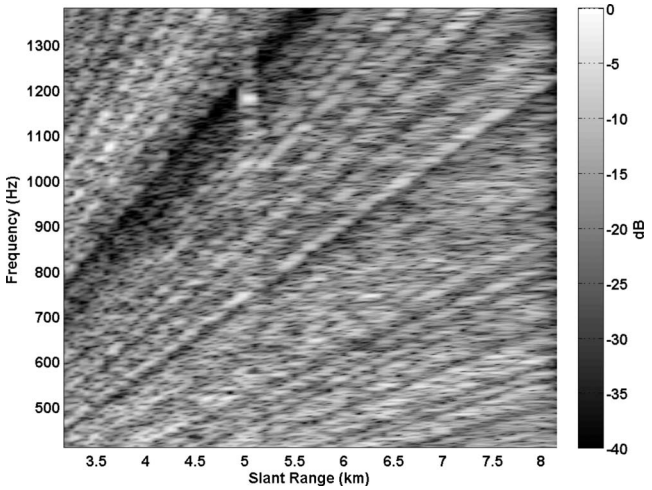


FIG. 2. STFT of a simulated reverberation return from a 400–1400 Hz signal. The weak target at 5.1 km is more visible than in the time-domain return, especially in the reverberation nulls around 1175 Hz.

III. WAVEGUIDE INVARIANT MODELING OF REVERBERATION

This section outlines the waveguide invariant principle and its application to modeling the time-frequency power spectra of reverberation returns. A consequence of multipath propagation is interference between propagating modes. For passive sonar applications, in the case of a moving broadband source, this modal interference results in structured patterns which can be observed in the time-frequency spectrum of the received signal.[17] Moreover, it has been found that these intensity patterns are well modeled over a range of environmental parameters as a function of source range, r, and frequency component, ω, relative to some reference range, r_0, and frequency ω_0,[21]

$$\omega=\omega_0\left(\frac{r}{r_0}\right)^\beta. \quad (7)$$

For a constant velocity source, Eq. (7) describes striations of constant intensity in the neighborhood of (r_0,ω_0). For range-independent environments, the waveguide invariant β is nearly constant, and defined as[17]

$$\beta=-\frac{\dfrac{\Delta k_{mn}}{\omega}}{\dfrac{\partial\Delta k_{mn}}{\partial\omega}}=-\frac{\dfrac{1}{v_m}-\dfrac{1}{v_n}}{\dfrac{1}{c_m}-\dfrac{1}{c_n}}, \quad (8)$$

where v_m and c_m are the phase and group velocities of mode m, respectively. This constant parameterizes the time-frequency interference pattern of many shallow water waveguides.

In this paper, an alternative formulation of the waveguide invariant given by Grachev[19] is used

$$k_m(\omega)-k_n(\omega)=\gamma_{mn}\omega^{-1/\beta}. \quad (9)$$

Here, γ_{mn} is a constant with respect to frequency, and the wavenumber difference for any two modes is related to frequency only through β. This will be used to derive a relationship between reverberation returns from two different range cells. For most range-independent, shallow-water waveguides, β has a value of approximately unity in both simulated and real data.[17] Rouseff shows that the value of β may be altered by mode coupling induced by internal waves and that source/receiver depth can be important in environments with large changes in the sound speed profile.[20] D'Spain and Kuperman discuss the impact of changes in bathymetry to the waveguide invariant. In significantly range-dependent waveguides, the waveguide invariant will also be range-dependent.[18] This work assumes a value $\beta=1$. Direct calculation of β is possible for a particular environment using the derivative of the group and phase slownesses as in Eq. (8), or by examining the frequency dependence of the wavenumber differences as in Eq. (9). Grachev's formulation of the waveguide invariant in Eq. (9) is used along with the data model presented in the previous section to derive a similar relationship governing the striations in the reverberation STFT. The details of the derivation can be found in the Appendix, however, the result is

$$\frac{E[|y(\omega,\tau,T)|^2]}{|U(\omega)|^2} \approx \frac{\alpha^{2/\beta+1}E[|y(\alpha^\beta\omega,\alpha\tau,\alpha T)|^2]}{|U(\alpha^\beta\omega)|^2}. \tag{10}$$

This relationship in Eq. (10) states that the normalized expected power of the reverberation return at a STFT bin of size T and centered at τ is equal to the normalized expected power from a different STFT bin of duration αT and center $\alpha\tau$ to a multiplicative constant α if the appropriate changes in frequency are made. This is the motivation behind the WI-CFAR detector and waveguide invariant-based reverberation estimate.

IV. WAVEGUIDE INVARIANT CFAR DETECTION

While conventional cell average techniques involve the normalization of the *A*-scan, e.g., in Fig. 1, waveguide invariant-based detection is based on the time-frequency distribution of the reverberation data, e.g., in Fig. 2. The test vector is defined as the STFT coefficients at the hypothesized target delay τ_0, $\mathbf{x}=[x(\omega_1,\tau_0,T),\ldots,x(\omega_N,\tau_0,T)]^T$. Detection of a target signal $\mathbf{s}=[s(\omega_1,\tau_0,T),\ldots,s(\omega_N,\tau_0,T)]^T$ in the presence of reverberation $\mathbf{y}=[y(\omega_1,\tau_0,T),\ldots,y(\omega_N,\tau_0,T)]^T$, from Eq. (6), and noise $\mathbf{n}$ can be posed as a binary hypothesis testing problem where

$$H_0\colon \mathbf{x}=\mathbf{y}+\mathbf{n}\sim CN(0,\mathbf{R}_c),$$

$$H_1\colon \mathbf{x}=\mathbf{s}+\mathbf{y}+\mathbf{n}\sim CN(\mathbf{s},\mathbf{R}_c). \tag{11}$$

The target return $\mathbf{s}$ in Eq. (11) is modeled as an unknown deterministic vector in reverberation-plus-complex Gaussian noise with covariance matrix $\mathbf{R}_c$. Assuming the STFT window length T is longer than the correlation length of the time series, the STFT coefficients of the return can be considered approximately uncorrelated. The covariance matrix $\mathbf{R}_c$ is thus a diagonal matrix with nonuniform variances $\sigma_k^2=\sigma^2(\omega_k,\tau_0,T)$, $k=1,\ldots,N$ on the diagonal.

If we temporarily assume that the coefficient variances σ_k^2 are known *a priori*, the generalized likelihood ratio test for Eq. (11) can be written as

$$\Lambda(\mathbf{x})=\frac{\prod_{i=1}^N p(x_i|H_1)}{\prod_{i=1}^N p(x_i|H_0)} = \frac{\max_s\left(\prod_{i=1}^N(\pi\sigma_i^2)^{-1}\exp\left(-\frac{|x_i-s_i|^2}{\sigma_i^2}\right)\right)}{\prod_{i=1}^N(\pi\sigma_i^2)^{-1}\exp\left(-\frac{|x_i|^2}{\sigma_i^2}\right)}, \tag{12}$$

where s_i and x_i are the ith element of $\mathbf{s}$ and $\mathbf{x}$, respectively. The maximum likelihood estimate of the signal $\hat{\mathbf{s}}$ can be obtained by maximizing the numerator of Eq. (12) with respect to $\mathbf{s}$, which yields $\hat{s}_i=x_i$. Using this estimate and taking the logarithm, the GLRT detection statistic simplifies to

$$\ln\Lambda(\mathbf{x})=\sum_{i-1}^N\frac{|x_i|^2}{\sigma_i^2}. \tag{13}$$

The detection statistic is the power in each STFT bin normalized by the reverberation-plus-noise variance and summed over frequency. The detector in Eq. (13) is derived assuming an *a priori* knowledge of σ_k^2. However, this quantity is unknown and must be estimated. The performance of the detector in Eq. (13) depends heavily on the accuracy of the estimate of the reverberation (and noise) power.

An estimate of the reverberation level can be obtained by using the result in Eq. (10). In this work, the reverberation level is estimated from regions where the expected power is shown to be constant

$$\hat{\sigma_{\mathrm{WI}}}^2(\omega,\tau,T)=\frac{1}{|I|}\sum_{\Delta\tau_i\in I}\left|x\left(\omega\left(1+\frac{\Delta\tau_i}{\tau}\right)^\beta,\ \tau+\Delta\tau_i,T\left(1+\frac{\Delta\tau_i}{\tau}\right)\right)\right|^2. \tag{14}$$

This corresponds to averaging the return along the striations. The range index i is taken over the set I, which includes bins adjacent in time, but excludes the target delay and surrounding guard bins, and $|I|$ is the number of elements in I. The change in frequency is made as dictated by the waveguide invariant. The proposed detector is compared against split-window normalization of the time series in cell-average CFAR detection. A frequency-dependent estimate of the reverberation power is also calculated without using the waveguide invariant property

$$\hat{\sigma_{\mathrm{FI}}}^2(\omega,\tau,T)=\frac{1}{|I|}\sum_{\Delta\tau_i\in I}\left|x\left(\omega,\tau+\Delta\tau_i,T\left(1+\frac{\Delta\tau_i}{\tau}\right)\right)\right|^2. \tag{15}$$

The STFT bins adjacent in time are averaged while holding the frequency constant, here referred to as a "frequency invariant" (FI) reverberation estimate.

To quantitatively compare these reverberation estimates, the MSE is calculated in each case. Though this can be done in several ways, it is most illustrative to begin by writing the MSE as a sum of the estimate variance and bias-squared

$$\mathrm{MSE}(\hat{\sigma^2}|\sigma_i^2,\sigma_s^2)=E[(\hat{\sigma^2}-E[\hat{\sigma^2}])^2]+(E[\hat{\sigma^2}]-\sigma_s^2)^2. \tag{16}$$

Looking first at the variance of the estimate, and noting that $E[\hat{\sigma}^2]=(1/N)\Sigma_{i=1}^N\sigma_i^2$, where σ_i^2 are the variances of the data points x_i used in the estimate, the first term in the above expression simplifies to

$$\mathrm{Var}(\hat{\sigma^2}|\sigma_i^2)=E\left[\left(\frac{1}{N}\sum_{i=1}^N|x_i|^2-\frac{1}{N}\sum_{i=1}^N\sigma_i^2\right)^2\right] = E\left[\left(\frac{1}{N}\sum_{i=1}^N(|x_i|^2-\sigma_i^2)\right)^2\right].$$

Since we have assumed the data are statistically independent, this becomes

$$\mathrm{Var}(\hat{\sigma^2}|\sigma_i^2)=\frac{1}{N^2}\sum_{i=1}^N\mathrm{Var}(|x_i|^2)=\frac{1}{N^2}\sum_{i=1}^N\sigma_i^4. \tag{17}$$

Next looking at the second term in Eq. (16),

$$\mathrm{Bias}^2(\hat{\sigma^2}|\sigma_i^2,\sigma_s^2) = \left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 - \sigma_s^2\right)^2 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_i^2\sigma_j^2 - \frac{2\sigma_s^2}{N}\sum_{i=1}^{N}\sigma_i^2 + \sigma_s^4. \quad (18)$$

The mean of the data at the point of interest (and along the associated striation) is σ_s^2. Substituting the above into Eq. (16), the MSE expression becomes

$$\mathrm{MSE}(\hat{\sigma^2}|\sigma_i^2,\sigma_s^2) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^4 + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_i^2\sigma_j^2 - \frac{2\sigma_s^2}{N}\sum_{i=1}^{N}\sigma_i^2 + \sigma_s^4. \quad (19)$$

In the waveguide invariant estimate, all of the data points have the same expected power since they lie along a striation, and the same expected power as the data under test. Therefore, $\sigma_i^2 = \sigma_s^2 \forall i$, and the above expression can be simplified

$$\mathrm{MSE}_{\mathrm{FI}}(\hat{\sigma^2}|\sigma_i^2,\sigma_s^2) = \frac{1}{N^2}\sum_{i=1}^{N}\sigma_i^4 + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_i^2\sigma_j^2 - \frac{2\sigma_s^2}{N}\sum_{i=1}^{N}\sigma_i^2 + \sigma_s^4, \quad (20)$$

$$\mathrm{MSE}_{\mathrm{WI}}(\hat{\sigma^2}|\sigma_s^2) = \frac{\sigma_s^4}{N}. \quad (21)$$

The bias of the waveguide invariant estimate is zero; the mean square error is due entirely to the variance of the WI estimate. An additional component of the error is present in the FI case, as this is averaging random variables which are not identically distributed, which is not an unbiased estimate.

The expressions in Eqs. (20) and (21) are conditioned on σ_i^2 and σ_s^2. To obtain the overall MSE, let σ_i^2 and σ_s^2 be uniformly distributed. The endpoints of the distribution can be directly related to the reverberation-to-noise ratio (RNR). The minimum value of the reverberation power is the noise floor and the distribution is bounded above by the noise power plus the maximum reverberation power, $[\sigma_n^2, \sigma_n^2 + \max(\sigma_s^2)]$, or normalized by σ_n^2, $[1, 1+\mathrm{RNR}]$. The normalized MSE averaged over σ_i^2 as a function of σ_s^2 is plotted in Fig. 3 for a reverberation-to-noise ratio of 10 dB. The error in the waveguide invariant estimate is universally lower than the frequency invariant case. The frequency invariant does nearly as well when the data being estimated are near the mean of the σ_i^2, and poorly in both extremes. Further, taking the expected value of the MSE with respect to σ_s^2 yields the expression

$$\mathrm{MSE}_{\mathrm{FI}} = \frac{4+N}{12N}\sigma_{s_{\max}}^4 + \frac{1}{N}\sigma_n^2\sigma_{s_{\max}}^2 + \frac{1}{N}\sigma_n^4, \quad (22)$$

$$\mathrm{MSE}_{\mathrm{WI}} = \frac{1}{3N}(\sigma_{s_{\max}}^4 + 3\sigma_n^2\sigma_{s_{\max}}^2 + 3\sigma_n^4). \quad (23)$$

Dividing by σ_n^4,

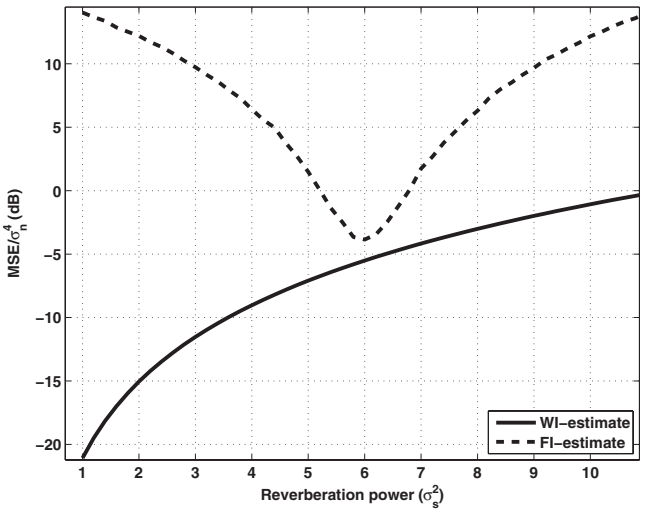


FIG. 3. Mean square error comparison of the FI and WI reverberation estimates, with 10 dB RNR.

$$\frac{\mathrm{MSE}_{\mathrm{EI}}}{\sigma_n^4} = \frac{4+N}{12N}\mathrm{RNR}^2 + \frac{1}{N}\mathrm{RNR} + \frac{1}{N}, \quad (24)$$

$$\frac{\mathrm{MSE}_{\mathrm{WI}}}{\sigma_n^4} = \frac{1}{3N}\mathrm{RNR}^2 + \frac{1}{N}\mathrm{RNR} + \frac{1}{N}. \quad (25)$$

The average MSEs in Eqs. (24) and (25) are only functions of RNR and the number of data points N used in the estimates. Figure 4 shows the effect of varying the RNR on the average MSE for both estimates. At very low RNR, the noise is dominant and using the waveguide invariant offers no benefit over averaging points selected randomly on the surface. When the reverberation is the principal term in the noise, striations are evident in the STFT and taking this structure into account leads to a 10 dB improvement in the MSE of the reverberation estimate.

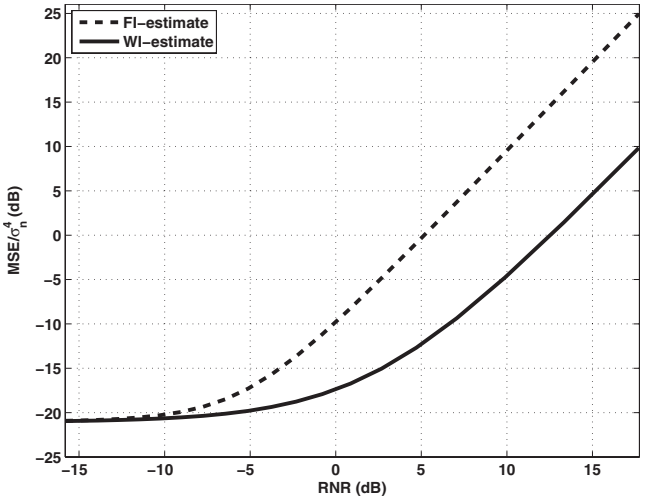


FIG. 4. Average MSE for the FI- and WI-estimates. At very low RNR, striations induced by the reverberation are not visible, and the WI-estimate provides no advantage.

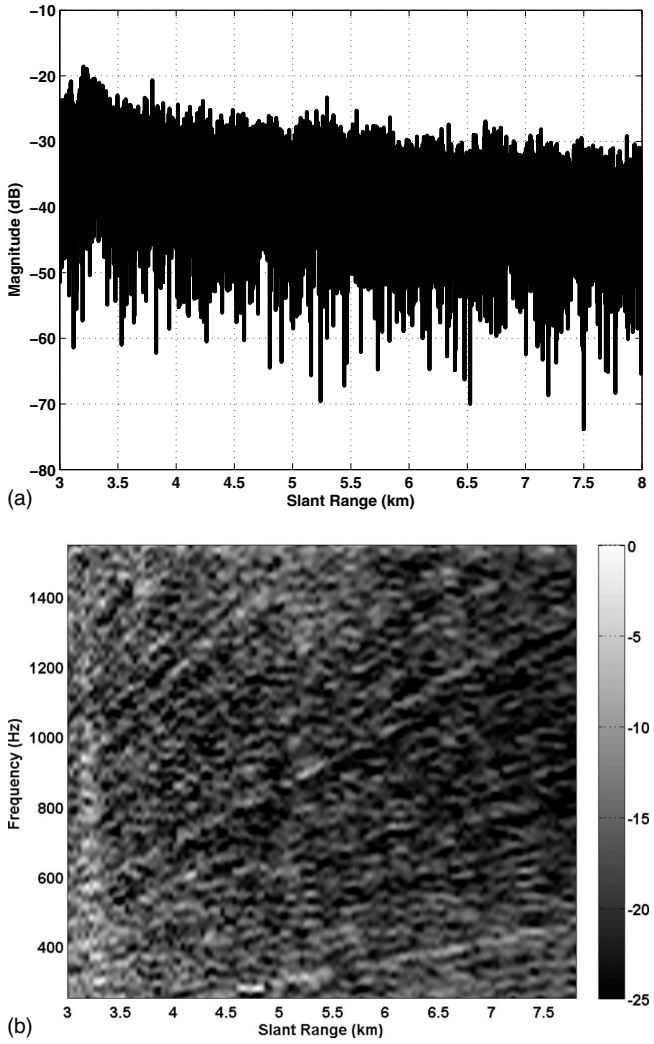


FIG. 5. Beamformed *A*-scan and STFT of beam 89 from the SCARAB98 experiment.

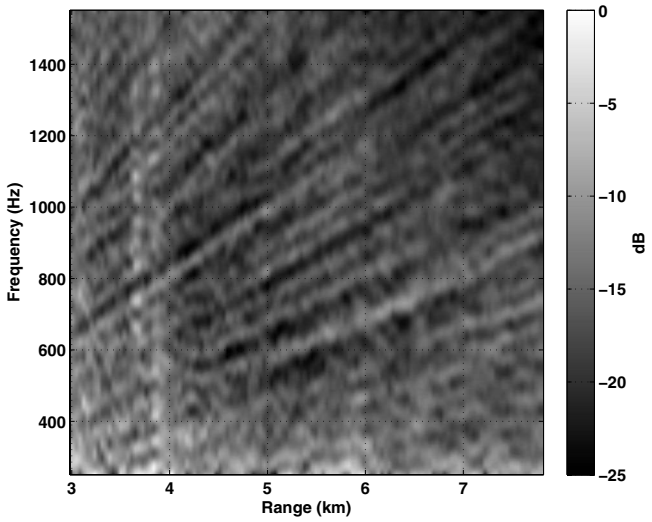


FIG. 6. Incoherent average of seven adjacent beams of data near endfire. The waveguide invariant striations are clearly visible. There is a strong return at approximately 4 km.

V. SIMULATION AND REAL DATA RESULTS

In previous work, detection based on the waveguide invariant was compared to conventional split-window normalization in a simulated Pekeris waveguide.[24] Here, the waveguide invariant method of reverberation estimation is applied to a more realistic scenario from the Malta Plateau region of the Mediterranean. The proposed detector is tested using both simulated and real data based on the SCARAB98 experiment.[16] The results are compared to both conventional cell-average techniques, and the frequency invariant method described above.

A. Real data results

Broadband shallow-water reverberation data were collected in the Malta Plateau region of the Mediterranean Sea off the coast of Sicily.[16] The data were collected on April 14, 1998 using a 128-element horizontal receiver array with 0.5 m element spacing. The water depth in the region varies from 130 to 150 m. The sound speed profile was slightly downward refracting, with a gradient of approximately 0.02 s^{-1}. Explosive sources [Mk61 Sound Undersea Signal (SUS)] were employed at a depth of 91 m. The sources were detonated several hundred meters from the array, however, a monostatic geometry was assumed since the ranges of interest are large. The reverberation return is sampled at 6 kHz and beamformed with Hanning shading into 90 beams each separated by approximately 2 deg in azimuth. The available bandwidth is about 100–1800 Hz; frequencies higher than 1800 Hz were attenuated with a low pass filter with this cutoff frequency. Further details of the environment and experiment can be found in Ref. 16. The reverberation return between 3 and 8 km from a single beam is shown in Fig. 5. Note that the STFT of the data shows the striation pattern predicted by waveguide invariant theory.

In order to compute reverberation level estimates, the STFT of the beamformed time series was calculated between slant ranges 3 and 8 km. The initial window length was 75 m and increased with slant range to accommodate the multipath spread, with 50% overlap. The window length is chosen to be much larger than the transmitted signal after pulse compression and multipath spread. The target is assumed to be completely contained in a single STFT bin, though spatially extended targets which extend into the guard bins will not affect the reverberation estimate. Figure 6 shows the incoherent average of seven beams of data, adjacent in azimuth and near endfire. The incoherent averaging reduces the noise and makes the reverberation more evident. The striation phenomenon is clearly illustrated in the Malta Plateau channel, and supports the hypothesis that $\beta \approx 1$ in this environment. There is a strong return at a slant range of 4 km that resembles a water column target. Ground truth for strong bottom features and scattering discretes were not available for these data, so it is unclear whether this is a sea-floor feature or a water column target.

Shown in Fig. 7(a) is a reverberation estimate calculated at (r, ω) according to the frequency-invariant method in Eq. (15). Care is taken not to use data in the vicinity of the test range r in the estimate. Excluding guard bands of approximately 100 m, 500 m of data in slant range was used on either side of the estimated range r, for a total of 1 km of

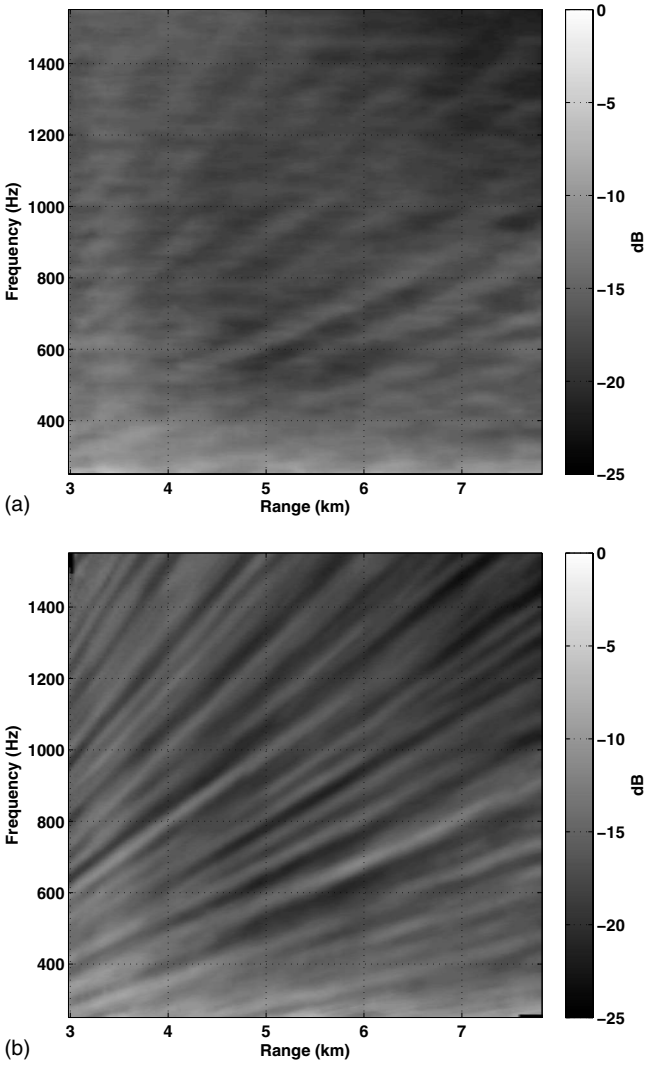


FIG. 7. Frequency (a) and waveguide invariant (b) estimates of the reverberation. Only the WI estimate retains the striated structure of the reverberation surface.

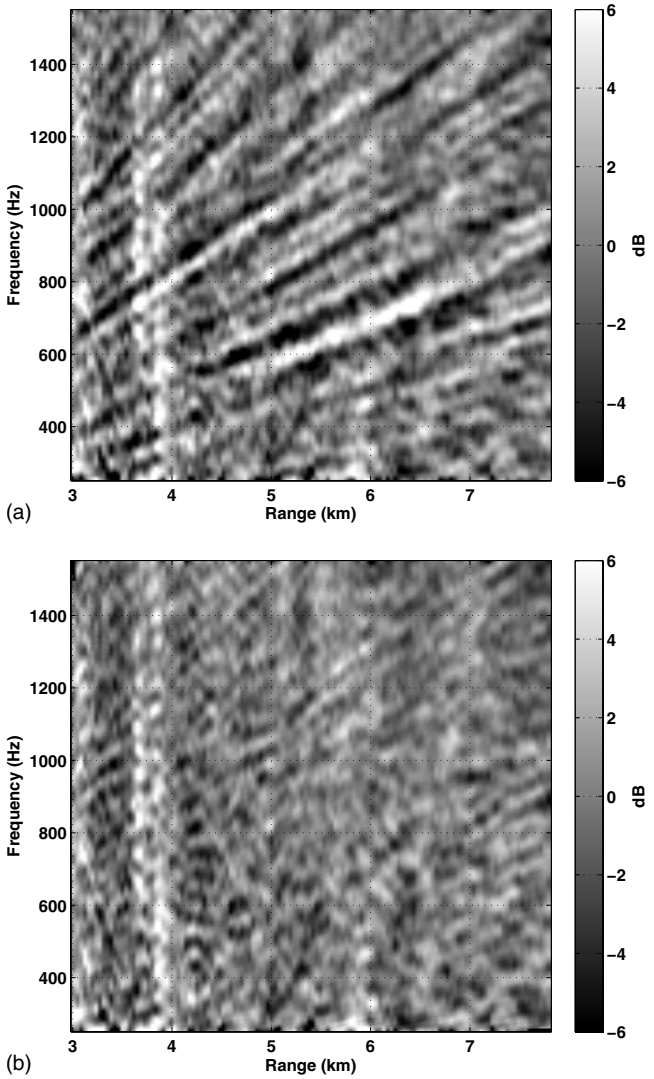


FIG. 8. Difference (in dB) between the data surface and the frequency invariant (a) and waveguide invariant (b) estimates.

data used in the estimates. Approximately 1000 Hz bandwidth was used in the simulated and real data reverberation estimates. Computing the waveguide invariant estimate averages along the striations requires more bandwidth than the frequencies of interest. For instance, a bandwidth of 400–1400 Hz can be used to estimate the reverberation level at 5 km for frequencies between approximately 500 and 1200 Hz if 100 m guard bins and 500 m of data are used on either side. The extra bandwidth is necessary to average along the striations at the highest and lowest frequencies in the frequency band of interest. The target-like return at 4 km is suppressed in the estimate, as would be expected in a reverberation level estimate. However, in comparing this estimate to the averaged STFT it is apparent that the frequency invariant estimate does not accurately capture the striations of the reverberation return in Fig. 6.

The image in Fig. 7(b) shows the estimate obtained by the WI method derived above. Here, the reverberation return at (r, ω) is estimated from data adjacent in slant range, and with a change in frequency according to the waveguide invariant, and as described in Eq. (14). The strong return at 4 km is similarly suppressed in this estimate. However, unlike the frequency invariant estimate, the striated nature of the original time-frequency surface is more accurately reproduced.

A further comparison of the two estimates is shown in Fig. 8. The surface in Fig. 8(a) shows the difference between the true reverberation data and the frequency invariant estimate. The target-like return is prominent in the difference surface, however, the error surface also has a striated pattern since the estimate does not capture this feature of the data. Figure 8(b) shows the difference between the averaged STFT and the waveguide invariant estimate. Again, the strong return at 4 km stands out clearly, but here the deviations are more uniformly distributed throughout the surface. This suggests that the structure in the STFT due to reverberation is more fully removed using a waveguide invariant estimate.

B. Simulation results

The SCARAB98 data are useful to demonstrate the predicted interference structure in real reverberation data, and to compare reverberation estimation methods. In order to assess

detection performance, the oceanographic parameters of the SCARAB98 experiment were used to generate simulated data. Realizations of these data were used to test the waveguide invariant detection scheme derived above against conventional CA-CFAR detection and FI-CFAR methods. The sound speed profile of the water column measured in the SCARAB98 experiment was used in the simulated data along with an estimate of the bottom composition. A water column depth of 127 m was used in generating the simulated data. The water column sound speed was slightly downward refracting, approximately 1515 m/s at the surface and 1511 m/s at the bottom interface. A bottom layer 2.5 m in depth with a 1580 m/s sound speed lay atop an acoustoelastic halfspace where the speed of sound was 1600 m/s. The eigenfunctions and horizontal wavenumbers were calculated using the KRAKEN normal mode program,[25] and simulated reverberation returns were calculated using the data model described above. The value of the waveguide invariant was verified for this environment by direct calculation according to Eq. (8). The waveguide invariant was found to be approximately unity for the simulated environment for all but the highest attenuating modes. White Gaussian noise was added to the simulated reverberation at a RNR of 10 dB to simulate the receiver noise in the SCARAB98 data. To generate data under the H_1 hypothesis, the simulated return from a point scatterer in the water column was added to the reverberation return to simulate a target. The source depth was 25 m. A simulated reverberation return was generated using a source depth of 75 m, with no significant change in the striation pattern. In the Mediterranean environment considered here, a change in the source depth has little effect on the striations in the reverberation since the water column is nearly isovelocity. When a thermocline is present, the source depth can affect the value of the waveguide invariant for certain groups of modes.[20] The strength of the target return was determined by the signal-to-reverberation ratio (SRR), defined as the peak target power to average reverberation power. The parameters of the STFT are the same as used above. Simulations were undertaken to compare the detection performance of conventional cell-average CFAR detectors, and the frequency-dependent detector derived above using frequency invariant and waveguide invariant estimates of the reverberation level. Figure 9(a) shows a comparison of the receiver operating characteristic (ROC) curves of each detector for the binary hypothesis problem described in Eq. (11). The range under test was 5 km and the SRR was set at 5 dB. The RNR remained fixed for all trials at 10 dB. The waveguide invariant detector shows significant improvement over both cell-average and frequency invariant methods. For a false alarm probability of $P_{FA}=0.1$, the waveguide invariant offers approximately double the detection probability of the other two methods. Additionally, the performance of the frequency invariant detector is comparable to the cell-average detector on the entire curve, suggesting that a frequency-dependent reverberation estimate is no more useful than a time-domain average.

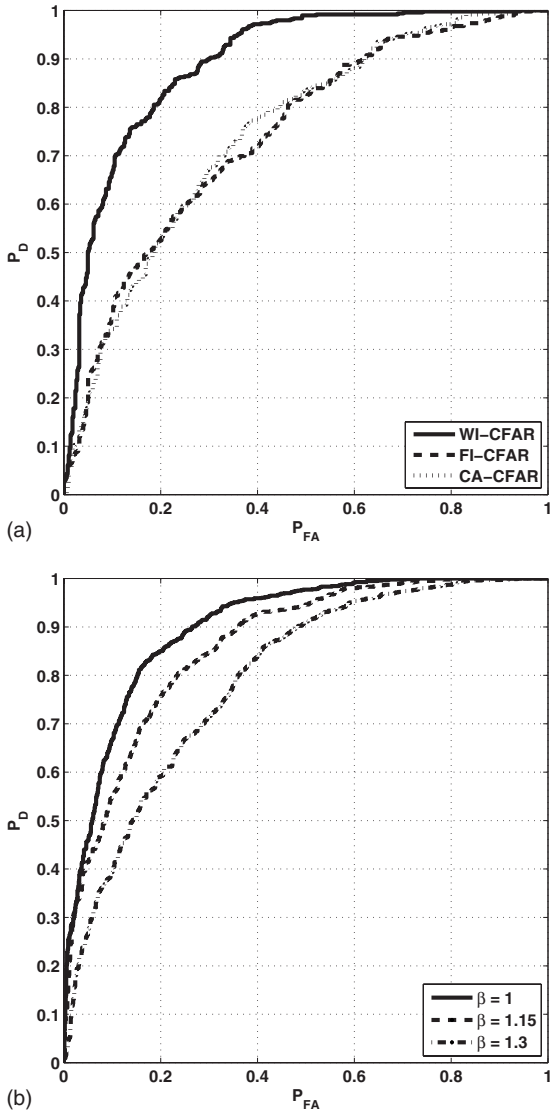


FIG. 9. ROC curves for CA-, FI-, and WI-CFAR detectors with SRR of 5 dB in the simulated SCARAB98 environment (a) and for the WI-CFAR detector using different values of the waveguide invariant (b). The value observed in the data is $\beta=1$.

C. Discussion

In this work, the waveguide invariant is assumed to be a known constant, $\beta=1$. Yet as previously mentioned, factors such as the sound speed profile, source depth,[20] and range-dependent bathymetry[18] can have an effect on the value of the waveguide invariant. Figure 9(b) examines the effect of a mismatch between the expected and observed value of β on detection performance. The detection performance with respect to assumed waveguide invariant values of 1, 1.15, and 1.3 are shown for the true case of $\beta=1$. For small mismatches in β ($<15\%$), the performance is nearly the same as the matched case. Using a predicted value of $\beta=1.3$, or a 30% mismatch from the true value, WI-CFAR detection offers no advantage over FI- and CA-CFAR methods in this environment.

The effect of SRR on detection performance is examined in Fig. 10. The probability of false alarm is fixed at 0.05, and the probability of detection is plotted against the SRR for all three detectors. Each point on the curves is the

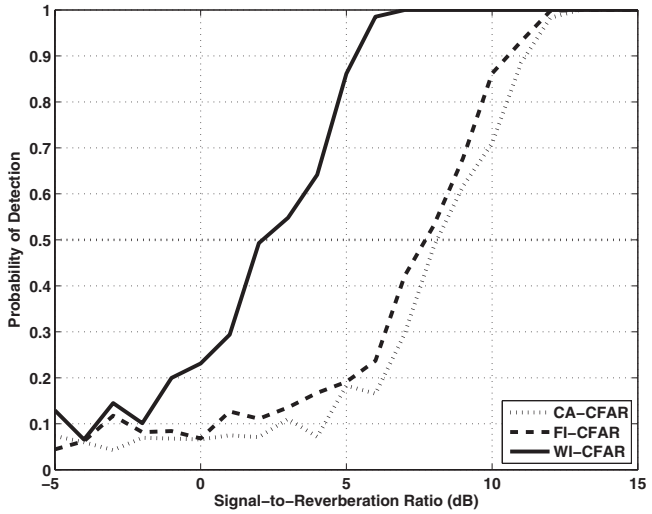


FIG. 10. Detection probability vs SRR for the CA-, FI-, and WI-CFAR detectors for P_{FA} fixed at 0.05. The WI-CFAR detector offers a gain of approximately 4 dB at $P_D=0.5$.

result of a 200 trial Monte Carlo simulation using the parameters described above. The waveguide invariant method offers a performance gain of approximately 4 dB at $P_D=0.5$.

VI. CONCLUSION

Many times, the STFTs of shallow-water reverberation returns have been seen to exhibit the striation phenomenon described by the waveguide invariant. This work has used these striations and the waveguide invariant property to estimate the frequency- and range-dependent reverberation level at the hypothesized target range. Results are presented for both simulated and real data, where the reverberation estimates based on the waveguide invariant are better able to capture the time-frequency structure of the data. Using a simple but suboptimal GLRT statistic, the waveguide invariant-based reverberation estimate improves detection performance. The performance of the new detector is evaluated for different reverberation levels and predicted values of β. The results presented here show good performance in a real and simulated range-independent and nearly isovelocity waveguide, when RNR is high and the waveguide invariant is known. Performance degrades when striations are less prominent (i.e., low RNR) or with a mismatch between the expected and true value of the waveguide invariant. The possibility of adaptively estimating the value of the waveguide invariant from neighboring reverberation data should also be explored. Applying prior knowledge of the sound speed profile and bathymetry may yield a more accurate value of β, and estimating β from surrounding reverberation data may provide environmental information about the waveguide.

ACKNOWLEDGMENTS

This work was supported by ONR Code 321US under Grant No. N00140510023. The authors would like to thank Charles Holland for providing the SCARAB98 data and Kevin LePage for many helpful discussions.

APPENDIX: WAVEGUIDE INVARIANT REVERBERATION MODELING

To model the time-frequency spectrum of broadband reverberation returns, the waveguide invariant relation in Eq. (9) can be applied to the STFT data model in Eq. (6). Let $W(\omega)$ and $g(\omega,r)$ be the Fourier transforms of $w(t)$ and $p(t,r)$ respectively. Then by the Fourier identity $\int g_1(t)g_2(t)dt=(1/2\pi)\int G_1(-\omega)G_2(\omega)d\omega$,

$$y(\omega,\tau,T)=\frac{1}{2\pi}\int_0^\infty\int_{-\infty}^\infty |T|W(-T\omega')\times g(\omega'+\omega,r)\exp(j\omega'\tau)\mathrm{d}\omega'\mathrm{d}r. \tag{A1}$$

From the normal mode propagation model

$$g(\omega'+\omega,r)=U(\omega'+\omega)\sum_{m,n}\eta(r)A_{mn}\frac{\exp(-j(k_m+k_n)r)}{r\omega}. \tag{A2}$$

The frequency dependence of the wavenumbers in the amplitude term have been approximated as $1/\sqrt{k_mk_n}\approx\kappa_{mn}/\omega$, where the frequency-independent constants κ_{mn} are lumped into A_{mn}. Using a first-order Taylor series expansion of the exponent about ω,

$$k_m+k_n\approx[k_m(\omega)+k_n(\omega)]+\left[\frac{\partial k_m}{\partial\omega}+\frac{\partial k_n}{\partial\omega}\right]\omega'. \tag{A3}$$

The partial derivative in the above expression is the reciprocal group velocity

$$\frac{\partial k_n}{\partial\omega}=\frac{1}{c_n} \tag{A4}$$

and so

$$g(\omega'+\omega,r)\approx U(\omega'+\omega)\sum_{m,n}\eta(r)A_{mn}\frac{\exp(-j(k_m+k_n)r)}{r\omega}\times\exp\left(-j\omega' r\left(\frac{1}{c_m}+\frac{1}{c_n}\right)\right). \tag{A5}$$

Assuming $U(\omega)$ is flat over the bandwidth and substituting Eq. (A5) in Eq. (A1), the inner integral in Eq. (A1) can be approximated as

$$U(\omega)\sum_{m,n}\eta(r)A_{mn}\frac{\exp(-j(k_m+k_n)r)}{r\omega}\int_{-\infty}^\infty |T|W(-T\omega')\times\exp\left(j\omega'\left(\tau-r\left(\frac{1}{c_m}+\frac{1}{c_n}\right)\right)\right)\mathrm{d}\omega'$$
$$\approx U(\omega)\sum_{m,n}\eta(r)A_{mn}\frac{\exp(-j(k_m+k_n)r)}{r\omega}w\left(\frac{r\left(\frac{1}{c_m}+\frac{1}{c_n}\right)-\tau}{T}\right). \tag{A6}$$

The window function $w(t/T)$ is nonzero over $[-T/2,T/2]$, therefore each term in the summation of the right-hand side of Eq. (A6) is nonzero when $-T/2<r(1/c_m+1/c_n)-\tau<T/2$. Letting $a_{mn}(\tau,T)=(\tau-T/2)(1/c_m+1/c_n)^{-1}$ and $b_{mn}(\tau,T)=(\tau+T/2)(1/c_m+1/c_n)^{-1}$, the interval is $a_{mn}(\tau,T)$

$< r < b_{mn}(\tau,T)$. Substituting into Eq. (A6) and putting the result into Eq. (A1) gives

$$y(\omega,\tau,T) \approx \sum_{m,n} \int_{a_{mn}(\tau,T)}^{b_{mn}(\tau,T)} \eta(r) U(\omega) A_{mn} \times \frac{\exp(-j(k_m + k_n)r)}{r\omega} \mathrm{d}r. \quad \text{(A7)}$$

The reverberation STFT of a time window centered at τ and of duration T can thus be approximated by summing the Fourier domain returns from point scatterers between ranges $a_{mn}(\tau,T)$ and $b_{mn}(\tau,T)$ for each mode combination (m,n). The intensity of the pressure field is

$$|y(\omega,\tau,T)|^2 \approx \sum_{m,n} \int_{a_{mn}(\tau,T)}^{b_{mn}(\tau,T)} \eta(r) U(\omega) A_{mn} \times \frac{\exp(-j(k_m + k_n)r)}{r\omega} \mathrm{d}r \times \sum_{m',n'} \int_{a_{m'n'}(\tau,T)}^{b_{m'n'}(\tau,T)} \eta^*(r') U^*(\omega) A^*_{m'n'} \times \frac{\exp(j(k_{m'} + k_{n'})r')}{r'\omega} \mathrm{d}r'. \quad \text{(A8)}$$

Since the scattering process is uncorrelated, $E[\eta(r)\eta^*(r')] = \sigma_\eta^2 \delta(r - r')$, the expectation of (A8) is only nonzero when $r = r'$, simplifying the expression to

$$E[|y(\omega,\tau,T)|^2] \approx |U(\omega)|^2 \sum_{\substack{m,n \\ m',n'}} A_{mn} A_{m'n'} \times \int_{A_{mnm'n'(\tau,T)}}^{B_{mnm'n'(\tau,T)}} \frac{\sigma_\eta^2}{r^2\omega^2} \exp[-j((k_m - k_{m'}) + (k_n - k_{n'}))r] \mathrm{d}r. \quad \text{(A9)}$$

The limits of the integral reflect the overlap of the integrals in Eq. (A8) defined by the group velocities of modes m and n, and m' and n',

$$A_{mnm'n'}(\tau,T) = \max(a_{mn}(\tau,T), a_{m'n'}(\tau,T)),$$

$$B_{mnm'n'}(\tau,T) = \min(b_{mn}(\tau,T), b_{m'n'}(\tau,T)). \quad \text{(A10)}$$

Now using Eq. (9),

$$E[|y(\omega,\tau,T)|^2] \approx \sigma_\eta^2 |U(\omega)|^2 \times \sum_{\substack{m,n \\ m',n'}} A_{mn} A_{m'n'} \frac{1}{\omega^2} \int_{A_{mnm'n'(\tau,T)}}^{B_{mnm'n'(\tau,T)}} \times \frac{\exp(-j(\gamma_{mm'} + \gamma_{nn'})\omega^{-1/\beta} r)}{r^2} \mathrm{d}r. \quad \text{(A11)}$$

We define the integral in Eq. (A11) as

$$P_{mnm'n'}(\omega,\tau,T) = \int_{A_{mnm'n'(\tau,T)}}^{B_{mnm'n'(\tau,T)}} \frac{1}{\omega^2 r^2} \times \exp(-j\Gamma_{mnm'n'}\omega^{-1/\beta} r) \mathrm{d}r, \quad \text{(A12)}$$

where $\Gamma_{mnm'n'} = \gamma_{mn} + \gamma_{m'n'}$. Changing the variables $u = \omega^{-1/\beta} r$, Eq. (A12) becomes

$$P_{mnm'n'}(\omega,\tau,T) = \int_{\omega^{-1/\beta} A_{mnm'n'(\tau,T)}}^{\omega^{-1/\beta} B_{mnm'n'(\tau,T)}} \frac{1}{\omega^{2+2/\beta} u^2} \times \exp(-j\Gamma_{mnm'n'}u)\omega^{1/\beta} \mathrm{d}u = \omega^{-2-1/\beta} \int_{\omega^{-1/\beta} A_{mnm'n'(\tau,T)}}^{\omega^{-1/\beta} B_{mnm'n'(\tau,T)}} \frac{1}{u^2} \times \exp(-j\Gamma_{mnm'n'}u) \mathrm{d}u. \quad \text{(A13)}$$

The above substitution removes all dependence on ω, τ, and T from the integrand; ω, τ, and T appear only outside and in the limits of the integral. Now consider the STFT coefficient at, $\tau' = \alpha\tau$ and $T' = \alpha T$, where α is a positive, real constant. If a change in frequency is also made according to $\omega' = \alpha^\beta \omega$, Eq. (A13) becomes

$$P_{mnm'n'}(\alpha^B\omega, \alpha\tau, \alpha T) = \alpha^{-2/\beta - 1} \omega^{-2-1/\beta} \times \int_{\omega'^{-1/\beta} A_{mnm'n'(\tau',T')}}^{\omega'^{-1/\beta} B_{mnm'n'(\tau',T')}} \frac{1}{u^2} \times \exp(-j\Gamma_{mnm'n'}u) \mathrm{d}u. \quad \text{(A14)}$$

The limits of the integral do not change with this substitution, since $a_{mn}(\tau,T)$ and $b_{mn}(\tau,T)$ are linear functions of τ and T and scaling in this case does not change the selection of the min or max functions, $\omega'^{-1/\beta} A_{mnm'n'}(\tau',T') = (\alpha^\beta\omega)^{-1/\beta} A_{mnm'n'}(\alpha\tau,\alpha T) = \omega^{-1/\beta} A_{mnm'n'}(\tau,T)$. Therefore,

$$P_{mnm'n'}(\alpha^B\omega, \alpha\tau, \alpha T) = \alpha^{-2/\beta - 1} \omega^{-2-1/\beta} \times \int_{\omega^{-1/\beta} A_{mnm'n'(\tau,T)}}^{\omega^{-1/\beta} B_{mnm'n'(\tau,T)}} \frac{1}{u^2} \times \exp(-j\Gamma_{mnm'n'}u) \mathrm{d}u = \frac{1}{\alpha^{2/\beta+1}} P_{mnm'n'}(\omega,\tau,T). \quad \text{(A15)}$$

Since the remaining terms in Eq. (A11) do not depend on range or frequency, the above result relates to the expected power of the STFT coefficients

$$\frac{E[|y(\omega,\tau,T)|^2]}{|U(\omega)|^2} \approx \frac{\alpha^{2/\beta+1} E[|y(\alpha^\beta\omega, \alpha\tau, \alpha T)|^2]}{|U(\alpha^\beta\omega)|^2}. \quad \text{(A16)}$$

[1] K. LePage, "Bottom reverberation in shallow water: Coherent properties as a function of bandwidth, waveguide characteristics and scatterer distributions," J. Acoust. Soc. Am. **106**, 3240–3254 (1999).
[2] D. Abraham, "Statistical normalization of non-rayleigh reverberation," in OCEANS'97 MTS/IEEE Conference Proceedings, 1997, pp. 500–505.
[3] K. LePage, "Statistics of broad-band bottom reverberation predictions in shallow-water waveguides," IEEE J. Ocean. Eng. **29**, 330–346 (2004).
[4] D. Abraham and A. Lyons, "Simulation of non-rayleigh reverberation and clutter," IEEE J. Ocean. Eng. **29**, 347–362 (2004).
[5] D. Middleton, "New physical-statistical methods and models for clutter

and reverberation: The ka-distribution and related probability structures," IEEE J. Ocean. Eng. **24**, 261–284 (1999).
[6]D. Abraham and A. Lyons, "Novel physical interpretations of k-distributed reverberation," IEEE J. Ocean. Eng. **27**, 800–813 (2002).
[7]D. Abraham and A. Lyons, "Reverberation envelope statistics and their dependence on sonar bandwidth and scattering patch size," IEEE J. Ocean. Eng. **29**, 126–137 (2004).
[8]C. Tindle and G. Bold, "Improved ray calculations in shallow water," J. Acoust. Soc. Am. **70**, 813–819 (1981).
[9]F. DiNapoli and R. Deavenport, "Theoretical and numerical green's function solution in a plane layered medium," J. Acoust. Soc. Am. **67**, 92–105 (1980).
[10]K. D. LePage, "Spectral integral representations of monostatic backscattering from three-dimensional distributions of sediment volume," J. Acoust. Soc. Am. **113**, 789–799 (2003).
[11]H. Bucker and H. Morris, "Normal-mode reverberation in channels or ducts," J. Acoust. Soc. Am. **44**, 827–828 (1967).
[12]R. Zhang and G. Jin, "Normal mode theory of the average reverberation intensity in shallow water," J. Sound Vib. **2**, 215–223 (1987).
[13]D. Ellis, "A shallow-water normal-mode reverberation model," J. Acoust. Soc. Am. **97**, 2804–2814 (1995).
[14]F. Ingenito, "Scattering from an object in a stratified medium," J. Acoust. Soc. Am. **82**, 2051–2059 (1987).
[15]N. Makris and P. Ratila, "A unified model for reverberation and submerged object scattering in a stratified ocean waveguide," J. Acoust. Soc. Am. **109**, 909–941 (2001).
[16]C. Holland, "Mapping seabed variability: Rapid surveying of coastal regions," J. Acoust. Soc. Am. **119**, 1373–1387 (2006).
[17]L. Brekhovskikh and Y. Lysanov, *Fundamentals of Ocean Acoustics* (Springer, New York, 1991).
[18]G. D'Spain and W. Kuperman, "Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**, 2454–2468 (1999).
[19]G. Grachev, "Theory of acoustic field invariants in layered waveguides," Acoust. Phys. **39**, 33–35 (1993).
[20]D. Rouseff, "Effect of shallow water internal waves on ocean acoustic striation patterns," Waves Random Media **11**, 377–393 (2001).
[21]J. Quijano and L. Zurk, "Use of the invariance principle for target tracking in active sonar geometries," OCEANS 2006 Boston MTS/IEEE Conference, 2006.
[22]J. Quijano and L. Zurk, "Demonstration of the invariance principle for active sonar," J. Acoust. Soc. Am. **123**, 1329–1337 (2008).
[23]K. D. LePage, P. Neumann, and C. Holland, "Broad-band time domain modeling of sonar clutter in range dependent waveguides," OCEANS 2006 Boston MTS/IEEE Conference, 2006.
[24]R. Goldhahn, G. Hickman, and J. Krolik, "Waveguide invariant reverberation mitigation for active sonar," ICASSP 2007, 2007, 941–944.
[25]M. Porter, The KRAKEN Normal Mode Program, SACLANT Undersea Research Center.

An active acoustic tripwire for simultaneous detection and localization of multiple underwater intruders

Thomas Folegot,[a] Giovanna Martinelli, Piero Guerrini, and J. Mark Stevenson
NATO Undersea Research Centre, Viale San Bartolomeo 400, 19126 La Spezia, Italy

(Received 28 September 2007; revised 24 July 2008; accepted 28 July 2008)

An algorithm allowing simultaneous detection and localization of multiple submerged targets crossing an acoustic tripwire based on forward scattering is described and then evaluated based upon data collected at sea. This paper quantifies the agreement between the theoretical performance and the results obtained from processing data gathered at sea for crossings at several depths and ranges. Targets crossing the acoustic field produce shadows on each side of the barrier, for specific sensors and for specific acoustic paths. In post-processing, a model is invoked to associate expected paths with the observed shadows. This process allows triangulation of the target's position inside the acoustic field. Precise localization is achieved by taking advantage of the multipath propagation structure of the received signal, together with the diversity of the source and receiver locations. Environmental robustness is demonstrated using simulations and can be explained by the use of an array of sources spatially distributed through the water column.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977675]

PACS number(s): 43.30.Vh, 43.60.Fg [RCG] Pages: 2852–2860

I. INTRODUCTION

The genesis of the barrier concept[1] described herein was research undertaken at NATO Undersea Research Centre (NURC). That research sought to answer some of the scientific questions underlying the practicality of an underwater surveillance system that incorporates an active acoustic tripwire at the mouth of a harbor and used to detect quiet submerged objects upon entering the system's field of view. In conducting our experimental studies, we attempted to build upon the successes of a lineage of earlier investigators.

In a progression of groundbreaking research dating from the early 1980s, Russian investigators demonstrated successful detection of moving underwater targets using forward-scattering techniques and published the theoretical background physics involved.[2–4] Experiments that validated the theory were conducted in Russian lakes using foam plastic targets roughly the size of human beings.[5]

Concurrently, an interesting discovery in the 1990s was that time-reversal acoustics could be a stable process even when undertaken in the highly time-dependent variability of the oceanic propagation medium.[6] This discovery brought the seminal experimental work of Parvulescu[7] into complex underwater propagation environments and built upon applications outside of acoustics examined up to that point (e.g., applications to nonlinear optics in space). Further work[1,8] extended the applied state of the science to show the first detection results using forward scattering in a time reversal mirror at 3.5 kHz, using an echo-repeater as a target in a waveguide of about 100 m depth. Tesei *et al.*[9] then showed that the phenomenon was reproducible using straightforward numerical acoustic propagation modeling and a simple geoacoustic field model of a waveguide, providing a first predictive capability to the technique. Further work by Gaumond *et al.*,[10] Carin *et al.*,[11] Yang,[12] Chambers,[13] Chambers and Berryman,[14] Folegot *et al.*,[15,16] and Prada *et al.*[17] showed the possibility of achieving time reversal focusing in shallower water at higher frequencies.

Since 1996, a series of sea trials were conducted at NURC to experimentally test acoustic time reversal methods in the sea. Through a progression of experiments, the concept of underwater, acoustic time-reversal focusing in the sea was brought from an unverified hypothesis to reality. When the experiments began, it was completely unknown whether the time-varying nature of the oceanic environment would support sufficient temporal stability for the method to succeed. A 10-year progression of experiments showed that not only was this possible, but that various applications could be implemented using higher frequency sound (up to 19 kHz), including some interesting acoustic barrier applications.[18]

The first experimental investigation into the performance of an underwater acoustic barrier based on forward scattering was conducted in the 11–19 kHz frequency band. We set out to design and test an array of sources that spanned an 11 m-depth harbor entrance and used as a signal of interest the aberration caused by the acoustic forward scattering of a crossing object. The chosen range was representative of the width of a harbor entrance (165 m).

This paper aims to demonstrate that localization of multiple targets can also be achieved together with detection: targets crossing the acoustic field issuing from a single source will each produce a shadow. Using a beamforming technique, the shadow for each acoustic path is estimated separately. It is shown that the shadow is produced only for a specific source depth and a specific acoustic path. In post-processing, a model is invoked to associate the expected paths with the observed shadow. The essential information is that the target is located at the intersection of the shadowed acoustic paths. This process, similar to triangulation, gives

[a]Electronic mail: folegot@nurc.nato.int

 © 2008 Acoustical Society of America

the position (range and depth) of the targets inside the acoustic field. Localization together with detection of multiple targets will then be demonstrated experimentally for submerged objects crossings at several depths and ranges.

A significant difference in this method is that the principle is based on the shadow produced by the target and therefore relies on the forward scattered signal and not the backscattered signal. The quality of the triangulation is achieved by taking advantage of the multipath propagation structure of the received signal, together with the spatial diversity of the arrays of sources and receivers that ideally span the whole water column to ensure robustness and resolution.

Environmental variability usually strongly affects detection performance. The rapid change of water temperature versus depth, due to diurnal temperature cycle, sunny-cloudy changes, or any local climatologic effects is addressed using simulations for several sound speed distributions. In conventional sonar systems, depending on their location in the water column, the change of the temperature field dramatically modifies the propagation structure. It usually creates false alarms of detection, or, more critically, avoids any possibility of detection in certain areas. For the proposed system that uses a multiple source array, while the geometry of the acoustic net created by the acoustic paths can be sensitive to environmental conditions, the size of the mesh remains stable for detection and localization.

This paper finally shows that, although the design of the experiment was not optimized for the proposed algorithm, there is a good agreement between theoretical performance and the at-sea data results.

II. TARGET LOCALIZATION

The implementation consists of an array of transmitters (active acoustic transducers) and an array of receivers, typically omni-directional hydrophones. It provides the natural geometry for an active acoustic barrier, or tripwire, which exploits observed changes in the received field caused by the introduction of an object into the field of view. A pulsed signal is emitted sequentially, with a given time interval, by each element of the transmit array on one side of the barrier and the refracted and/or reflected signal is observed on the other side. To be effective, the time delay must be longer than the impulse response of the propagation medium. If the property of spatial reciprocity holds in the medium, i.e., if the forward and backward Green's functions are identical, then we can also extract signals as if they had been sent in the reverse direction, although physical sound signals are sent through the water in one direction only. Then it is expected that the presence of a target will produce a shadow on both sides of the system, for certain rays, and certain acoustic sensors. The shadow effect in conjunction with an acoustic model of the environment is shown to allow the identification of specific rays crossing the target position. Those particular rays are then used to triangulate the position of the target.

A. The acoustic shadow effect

An object's acoustic shadow is commonly used in mine-hunting sonar processing as an aid to object classification[19] in backscattering mode. The shadow in these applications is defined as the silhouette of an object, cast onto the seafloor behind the object.[20] For a forward-scattering barrier, the shadow created by a target is defined as the reduction of the energy transmitted from one side of the barrier to the other side. The reflection and diffraction of the acoustic energy in all directions except the nominal incident direction and the absorption of the acoustic energy by the target itself cause this decrease. It can be seen as the destructive interference between the incident field p^{inc} and the scattered field p^{scat}. Therefore, the total pressure received at a given receiver is a function of the source, receiver, and target position, and the eigenrays l between the source and the receiver crossing the target

$$p^{\mathrm{tot}} = p^{\mathrm{inc}} + \sum_{l} p_l^{\mathrm{scat}}. \tag{1}$$

The intrusion of a scatterer into the propagation medium results in a transient shadow that provides a signal of interest in an otherwise stable field observed at the receive array. Working with generally low target strengths, post-processing algorithms are applied in order to emphasize the signal excess and triangulate for the location of the object.

B. Description of the localization algorithm

We use a traditional delay-and-sum beamformer defined for plane-waves as[21]

$$B(\vec{\alpha}, \tau) = \sum_{h} w_h p_h(\tau + \vec{\alpha} \cdot \vec{r}_h), \tag{2}$$

where $p_h(\tau)$ is the signal received at time τ by hydrophone h located at $\vec{r}_h = x_h + jz_h$, and z_h and x_h are depth and range, respectively. The amplitude weighting w_h enhances the beam's shape and reduces sidelobe levels. The slowness vector is defined by $\vec{\alpha} = \vec{\zeta}/c$, where $\vec{\zeta}$ is a particular propagation direction, and c is the sound speed. When the source is located in the array's near field, the equivalent beamforming for spherical waves would be more appropriate.[21]

Consider that, for a given ping, $N_{s,l}$ represents the acoustic intensity measured by the beamformer when the direction of propagation $\vec{\alpha} = \vec{\alpha}_l$ matches the acoustic path l associated with source s. In practice, $|B_s(\vec{\alpha}_l, \tau)|^2$ is integrated across a short period of time Δt centered on the propagation time

$$N_{s,l} = \int_{-\Delta t/2}^{\Delta t/2} |B_s(\vec{\alpha}_l, \tau)|^2 d\tau. \tag{3}$$

Any target positioned along the acoustic travel path associated with $N_{s,l}$ will shadow the acoustic beam and produce a sudden energy decrease toward the ambient noise level. A target localization ambiguity image is created using the sudden decreases for a subset of $N_{s,l}$, where s describes all contributing source elements, and l all contributing acoustic paths. The proposed method uses the approximation of Gaussian ray bundles[22]

$$\Psi_{s,l}(\vec{r},T)=\left(\left(\frac{\langle N_{s,l}(t)\rangle_{t<T}}{N_{s,l}(T)}\right)^{\gamma}-1\right)\cdot e^{-(z-z_{s,l}(r))^2/2\sigma^2}, \quad (4)$$

in which $\vec{r}=(r,z)$ is a coordinate of the acoustic field, the width of the bundle σ is usually chosen as half of the wavelength, $z_{s,l}(r)$ describes the geometry of the ray emanated from the source s, and associated to path l, $\langle N_{s,l}(t)\rangle_{t<T}$ is the average energy level associated to source s and path l prior to the emission time T, and γ is an empirical contrast coefficient which allows an increase in the weight of the most attenuated paths in the reconstruction of the final image. $\Psi_{s,l}(\vec{r},T)$ corresponds to the weighted ray geometry l associated with source s, where the weight coefficient is the inverted loss observed on the received signal in direction $\vec{\zeta}_l$. A propagation model is then invoked to estimate $z_{s,l}(r)$ associated with $\vec{\zeta}_l$. For a given ping received at time T, the sum of all acoustic bundles $\Psi_{s,l}(\vec{r},T)$ for all acoustic elements and paths will produce an ambiguity image defined by

$$I(\vec{r},T)=\sum_s\sum_l \Psi_{s,l}(\vec{r},T). \quad (5)$$

Only shadowed travel paths with lower $N_{s,l}$ than the mean level (level obtained without target) will significantly contribute to the ambiguity image. When the energy is equal to the mean level, e.g., there is no perturbation along the (s,l) travel path, there is no contribution to the ambiguity image. The triangulation of all these travel paths yields the positions of the targets. The intersection of all paths $z_{s,l}(r)$ at the target location $\vec{r}_{\text{target}}$ will maximize $I(\vec{r})$,

$$I(\vec{r}_{\text{target}})=\max_{\vec{r}}(I(\vec{r})). \quad (6)$$

C. Theoretical performance

Theoretical performance is evaluated using a method that evokes standard ray theory. The method is based on the identification of all the ray geometries $z_{s,l}(r)$ which cross the target position. The loss of energy $N_{h,l}/\langle N_{h,l}\rangle$ associated with those particular rays is arbitrarily chosen to be 0.8, i.e., 20% of the incident acoustic energy is absorbed or diffracted by the target. The computed ambiguity image $I(\vec{r},T)$ is a theoretical estimation of the performance of the system.

1. Ideal coverage and resolution

An ideal hardware setup, i.e., one envisioned in an operational system installed in a harbor on permanent basis, might consist of a source array and a receiver array spanning the water column from the bottom to the surface at the harbor entrance. In the case of only four acoustic elements on each side of the harbor entrance, the top of Fig. 1 shows how the geometry of the system, in particular the spanning of the whole water column, is key to optimizing acoustic coverage of the entrance. The mesh of the acoustic web created by the multipath propagation from all sources to all receivers depends on the distribution of the acoustic transducers, and the number of bounces on the interfaces of the waveguide. The more reflection along the interfaces, the more paths intercept the target location and can then be used to localize it within the ensonified field (Fig. 2).

Let us consider for instance an acoustic tripwire working at 15 kHz central frequency, in which all acoustic elements are vertically separated by 1 m. Figure 3 shows the theoretical ambiguity image $I(\vec{r})$ for a 165 m width and 11.5 m deep harbor entrance, when one or two underwater intruders are crossing the acoustic field. The intruder radius is set to 0.25 m, and its associated acoustic loss to 20% of the incident energy. The detection and localization of the multiple targets is achieved with a good signal to noise ratio, due to the numerous rays crossing the target positions, as shown in Fig. 2. In fact, the signal to noise ratio of the obtained ambiguity image $I(\vec{r})$ (Fig. 3) is proportional to the ratio between the total number of paths that are intercepted by the target, and the overall number of ray intersections at other locations in the field. The quality of the shadow produced by the target depends on the ratio between the size of the target and the acoustic wavelength used by the system, and the material of the target. Any absorbing or diffracting material will be imaged by the system, from solid targets like scuba tanks and manned or unmanned submersibles, including wet suits. Detection performances are enhanced by an increased number of elements on each side of the tripwire, an increased number of reflections on the interfaces, and a greater spatial diversity of the acoustic elements.

2. Robustness to environmental changes

As harbors are designed to protect ships from winds and surface roughness, the primary environmental effect to be considered is rather the possibility of rapid changes in the water-column temperature profile due to the diurnal temperature cycle, solar radiation changes, or local climatologic effects. The test case described in the previous section is applied to (a) an isovelocity profile and (b) a two-layer density profile that might typically be encountered during a sunny afternoon. Although the propagation geometry of the ray has changed (Fig. 1), the combination of the diversity of source location in the water column and the exploitation of the reflections on the boundaries is key to maintaining a high level of acoustic coverage, and therefore the performance of the system.

The bottom avoids acoustic rays at incident angles higher than the associated critical angle. Thus, the density of the acoustic web will be impacted by a low critical angle, since the number of bottom reflected paths will be reduced. The bottom-roughness effect on the propagation is frequency dependent. The effect is primarily important at mid (i.e., 1–10 kHz) and high (i.e., >10 kHz) frequencies. At lower frequencies (i.e., <1 kHz), they are less important. At frequencies lower than 100 Hz, this bottom effect becomes almost insignificant. However, as the bottom roughness is invariant in time, it should not impact the change of $N_{s,l}$ for a given source s and acoustic path l, especially if the bandwidth of the signal is large.

On the other hand, the impedance difference between air and water ensures the total reflection of the acoustic waves, but a time-varying sea surface—especially the wakes produced by transiting surface ships—will certainly increase significantly the variance of $N_{s,l}(t)$, and therefore degrade the performance of the detection algorithm.

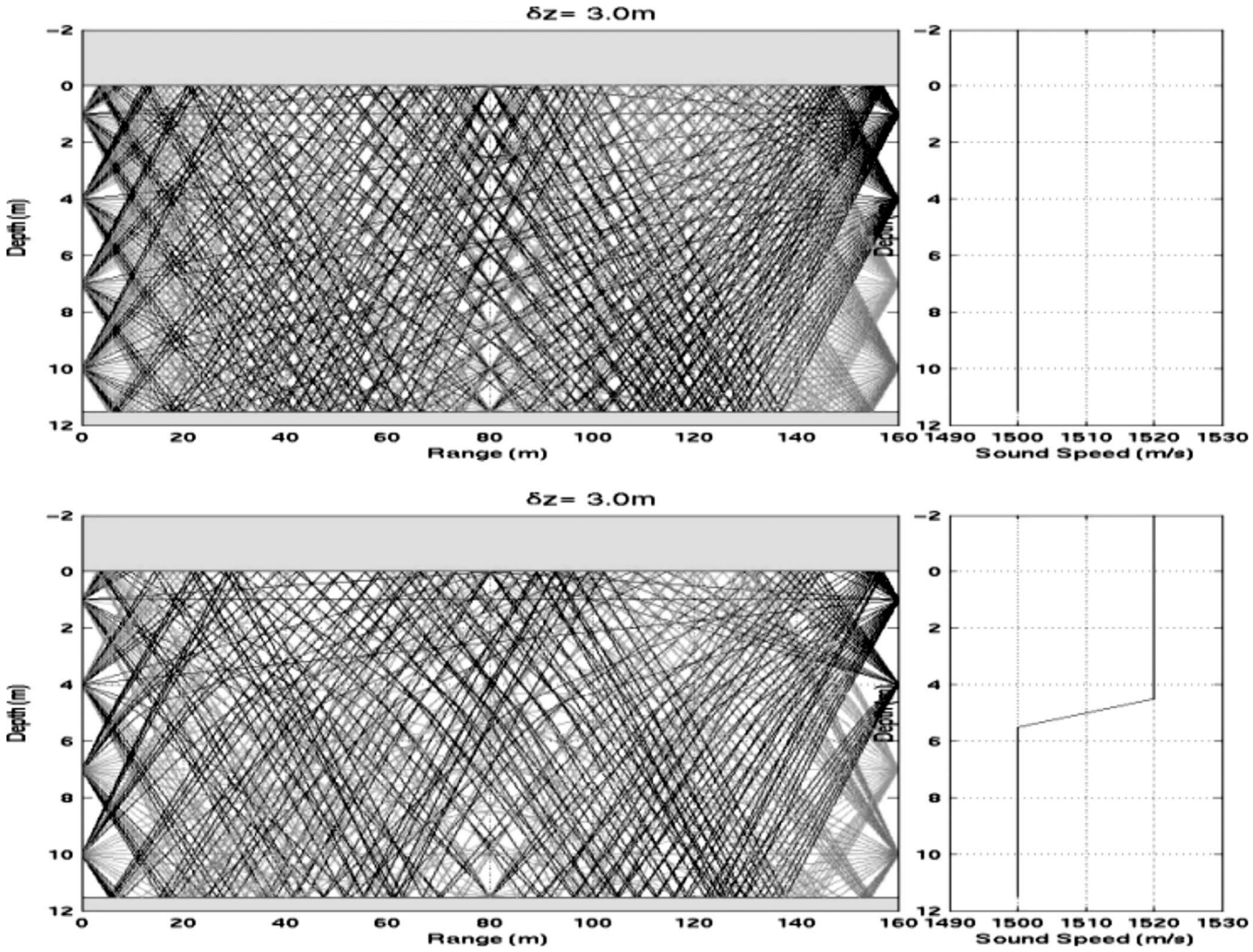


FIG. 1. Coverage of the acoustic tripwire for only four acoustic elements spanning the whole water column on each side of the harbor entrance, for isovelocity (top) and two-layer (bottom) environments. Although the propagation structure is modified, the mesh of the net remains unchanged.

III. DEMONSTRATION AT SEA

The experiment, called BARRIER-04, was conducted at a center frequency of 15 kHz in a shallow water waveguide with a hard seafloor interface. About 8 kHz of acoustic bandwidth was available. The waveguide length was typically about 16 water depths, with one of the waveguide interfaces being an undulating seafloor and the other being the time-varying sea surface. Moving objects (e.g., steel spheres) were towed through in the water column, while short pulses were emitted continuously from one to the other side of the barrier. A subset of the total acoustic data gathered during the sea-trial was suitable to demonstrate both the proposed concept and algorithm of multiple target detection and localization.

A. Geometry of the experiment

The experiment was conducted in October 2004 off Tellaro, Italy, in a shallow water waveguide harbor setting (Fig. 4). On the receive end of the tripwire, a linear vertical receive array (VRA) consisting of 16 omnidirectional hydrophones spanned 8.3 m, roughly half of the water column. The inter-element spacing between the hydrophones was either 0.36 or 0.72 m. The individual hydrophone sensitivity was −202.5 dB re 1 V/μPa up to the designed low-pass frequency roll off at 20 kHz. The hydrophone preamplifier gain was 25 dB. The voltage signals from the 16 channels were digitized at a sampling rate of 50 kHz with 16 bits resolution. For synchronization, GPS referenced clock time stamping was used.

A vertical array of acoustic transmitters (so-called "SRA") was placed on the other end, about 165 m away. The transmit array consisted of a rigid vertical tower of sources made from 12 Simrad transducer staves with inter-element spacing 0.595 m. The bottom-most transducer was located at 0.3 m from the seafloor. The beam width of the source transducers was 19 deg in the horizontal plane and 44 deg in the vertical plane. The maximum source level of any one transducer was 185 dB re 1 μPa at 1 m. The transducer could broadcast in an acoustic frequency range from roughly 11 to 19 kHz.

The at-sea experiment conduct was as follows. We transmitted pulses consisting of two cycles of a 15-kHz sine wave sequentially from all transducers in the source array. The time interval between two successive pulse transmissions was 50 ms. One complete sequence of twelve transmissions,

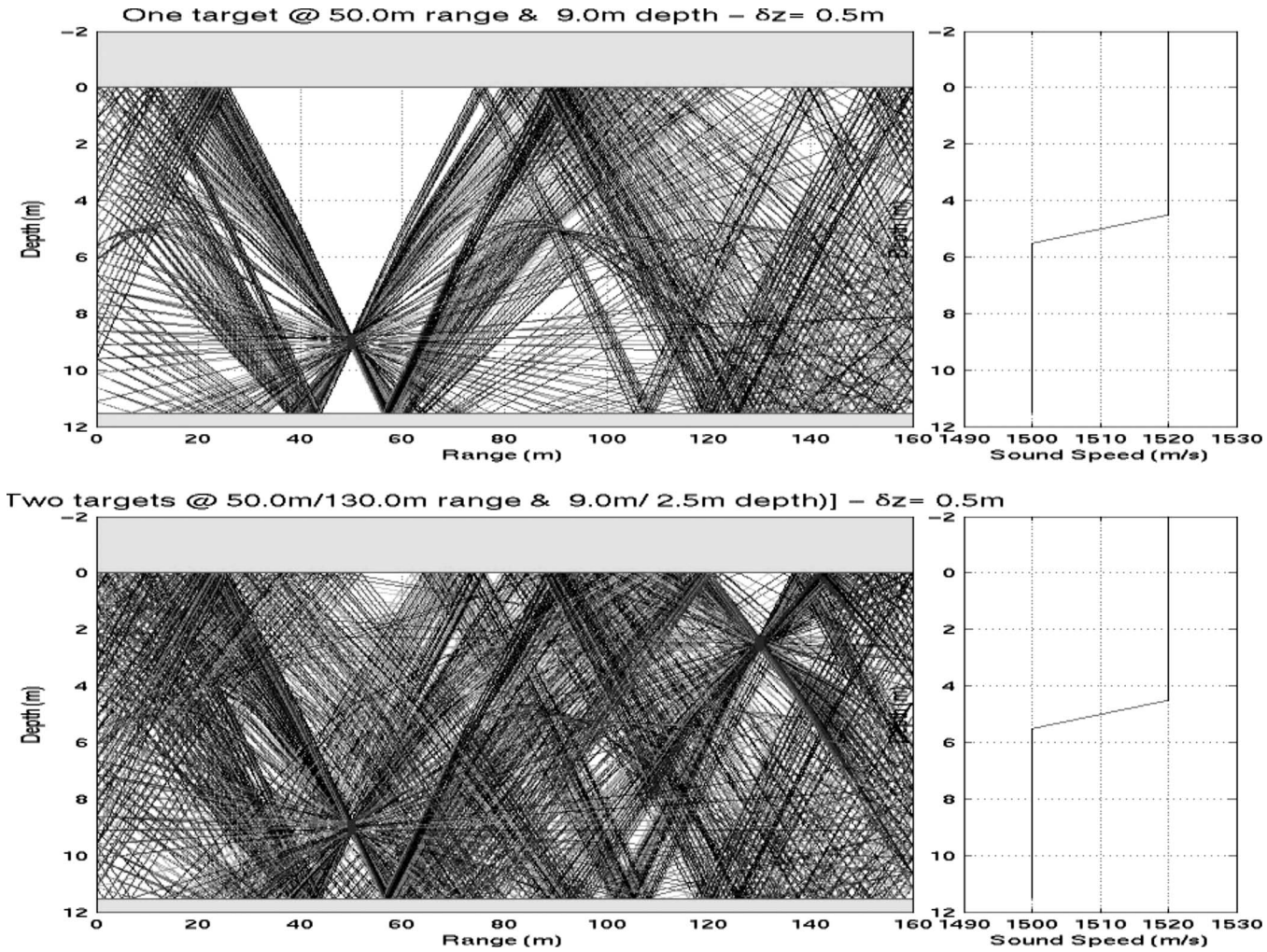


FIG. 2. Identification of the rays crossing a single target (top) and two targets (bottom). Both targets are 0.5 m diameter spheres, arbitrarily placed at 50 m range and 9 m depth, and at 130 m range and 2.5 m depth.

called a "round-robin" for naming convenience, was transmitted every second. Simulated intruders were introduced at several ranges and depths into the field by towing submerged objects (by a human being pulling a light line at speeds of the order of 0.25 m/s) through the vertical plane between the SRA and the VRA. An independent (visual) cue allowed verification of the object's barrier crossing time to within about ±5 s.

B. Coverage

Due to the limited number of transducers and hydrophones, the limited aperture of both arrays, and the relative wide spacing between the elements, beamforming was accomplished using all signals of either the 12-element source array, or the 16-element receiver array. Therefore, the acoustic web created by the system—considering a maximum of one bounce on each boundary—is nonuniform and sparse, as shown in Fig. 4. Notwithstanding these limitations, we applied the localization algorithm to this configuration to test in-water performance to the expected theoretical performance.

C. Experimental results

This section describes the signal processing applied to the data gathered during the experiment and analyzes the results in terms of stability of the received energy over time.

1. Data processing

The algorithm described in Sec. II B is applied to the acoustic data gathered the first day of the BARRIER-04 sea trial. Since the combined use of a source array and a receiver array allows measurement of all the transfer functions between all sensors across the channel, each array can be equally seen as a receiver array or as a source array. Using the pressure field measured successively along both arrays, $\{p_h\}_{h=1,N_{rcv}}$ and $\{p_s\}_{s=1,\ldots,N_{src}}$ for the $N_{rcv}=16$ elements of the VRA and the $N_{src}=12$ elements of the SRA, respectively, the beamforming is defined for each path l and l' between the center of phase of the arrays and the sources

$$B_s(\vec{\alpha}_l,\tau)=\sum_{h=1}^{h=N_{rcv}} w_h p_h(\tau+\vec{\alpha}_l\,.\,\vec{r}_h) \quad \text{for } s\in\{1,\ldots,N_{src}\}, \tag{7}$$

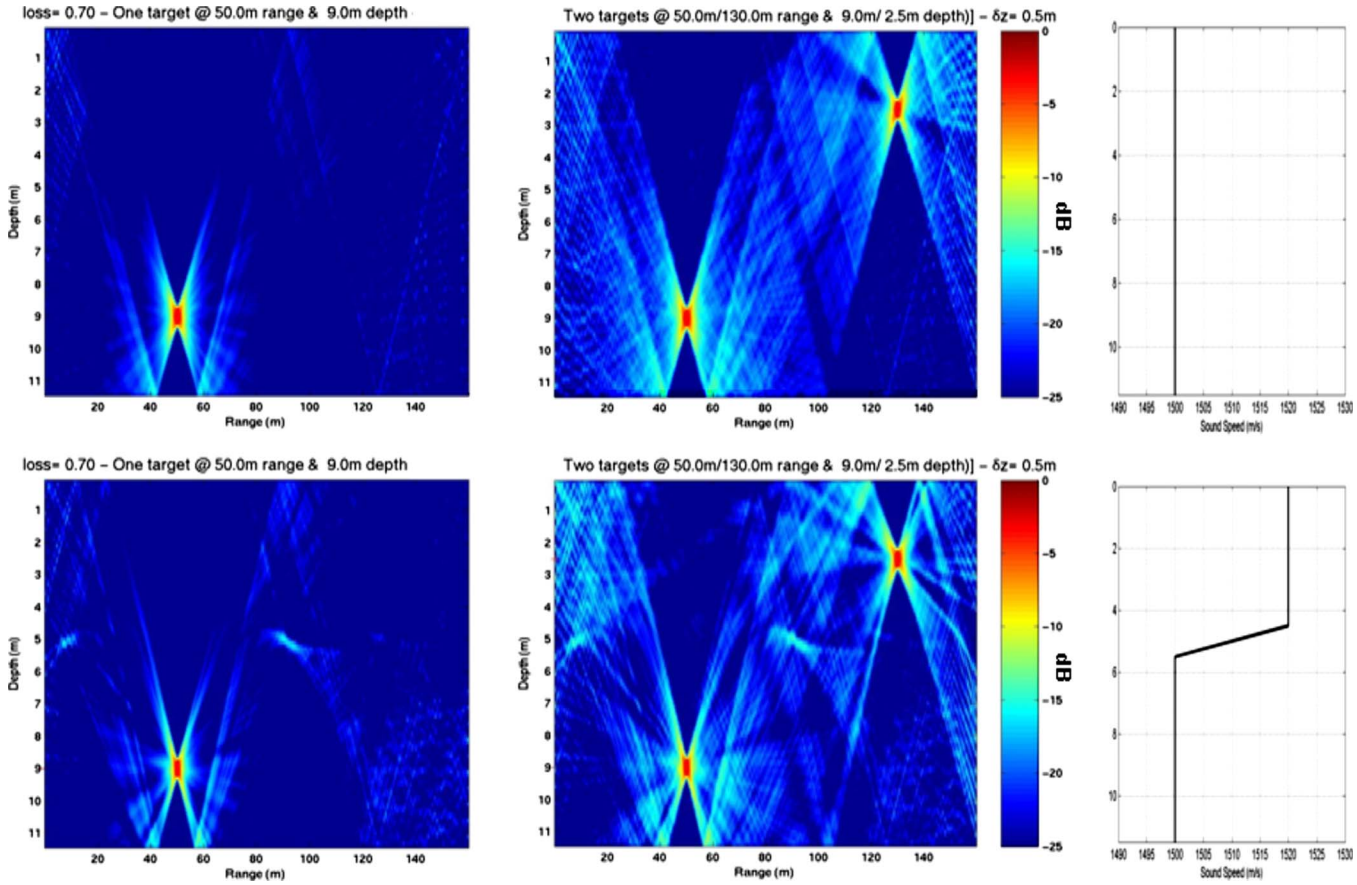


FIG. 3. (Color online) Simulated ambiguity images for one (left) and two (center) targets for an isovelocity (top) and a two-layer (bottom) environment. The shading coefficient is arbitrarily set to 20% for each ray, and the contrast coefficient γ to 4.

$$B_h(\vec{\alpha}_{l'}, \tau) = \sum_{s=1}^{N_{src}} w_s p_s(\tau + \vec{\alpha}_{l'} \cdot \vec{r}_s) \quad \text{for } h \in \{1, \ldots, N_{rcv}\}. \tag{8}$$

The indices l and l' indicate the ray geometry. The experimental data have been processed using the direct path, the first bottom bounce, the first surface bounce, the bottom and surface bounce, and the surface and bottom bounce. The integration time Δt used to estimate $N_{s,l}$ and $N_{h,l'}$ is equal to the length of the emitted signal.

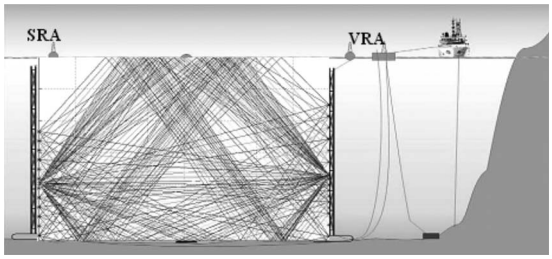


FIG. 4. The experiment configuration, not drawn to scale. The water depth is 11.5 m. The 12 transducer array (SRA) spans 6.8 m of the water column. The 16 hydrophone VRA spanned 8.5 m of the water column. The two arrays were cabled to a common data acquisition point aboard a ship. The coverage (or acoustic "net") of the tripwire is superimposed on the set-up, considering only the rays which have a total number of two bounces on the interfaces. The limited number of sources and receivers impacts the quality of the coverage.

2. Evidence of a shadowing effect

The estimation of $N_{s,l}$ and $N_{h,l'}$ is shown in Fig. 5 between 120 and 210 s. During this run, the crossing of the target occurs between 170 and 190 s. The figure shows that some shadows appear (as black spots) around the time of the crossing, but only for some specific paths and some specific sensors. For example, looking carefully at the energy received by the VRA (first column of Fig. 5), the energy clearly drops down to almost the noise level during a few seconds, simultaneously along:

(a) the direct path for the 10th element of the SRA,
(b) the bottom reflected path for the 8th element of the SRA, and
(c) the surface reflected path for the 6th element of the SRA, etc.

The same kind of shadowing is observed when the beamforming is done using the received pressure of the SRA (second column of Fig. 5).

3. Imaging the crossing targets

The geometrical correlation of acoustic shadows along some particular paths between the source array and the receiver array is now used to estimate the location of the target.

Figure 6 shows the results of the localization algorithm for three crossing configurations: a target crossing at 92 m

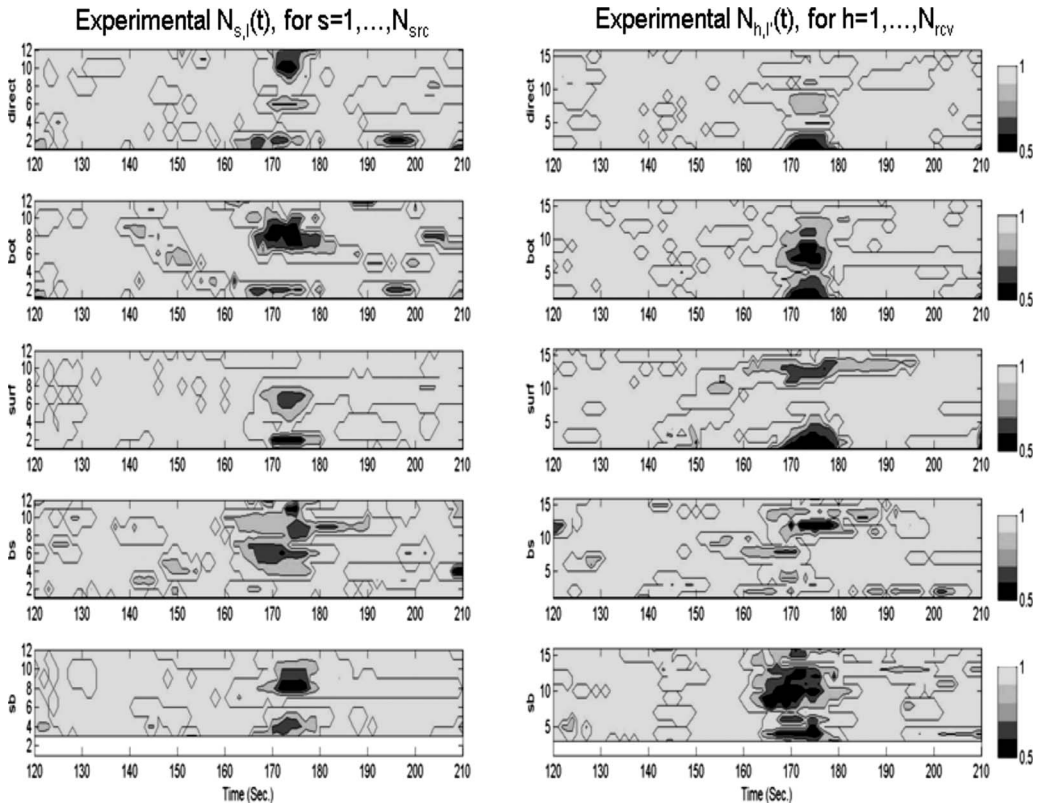


FIG. 5. Experimental losses observed during the BARRIER04 sea trial. The losses are displayed for the 12 elements s of the source array (left) and for the 16 elements h of the receiver array (right), for direct path, bottom ("bot"), surface ("surf"), bottom then surface ("bs") and surface then bottom ("sb") bounced rays. The target was crossing the tripwire around $T=180$ s. In each case, a loss of energy is clearly shown for some elements and some acoustic paths.

range and 6.8 m depth (first row), a target crossing at 45 m range and 9.5 m depth (second row), and a target crossing at 115 m range and 8.5 m depth (third row). All resulting images show a convergence of energy at the location of the target for all three crossings. However, in each case, another concentration of energy can be seen near the surface at the same range as the target. This is explained by the presence of the surface buoy used to maintain the target at a constant depth and manually pull it through the water. To address this, the figures also show the theoretical image (left column) obtained by simulating the same configuration as the experiment (experimental results displayed on the right column). To produce these theoretical images, an additional target was placed on the surface to model the air filled surface buoy.

The sparse nature of the acoustic web created in this particular experiment introduces ambiguities in the final images which do not correspond to any target location. The amount of paths effectively affected by the target is relatively limited, and of the same order as the number of intersections among these same paths at some other locations. This occurs for example in the top-right panel of Fig. 6 at depth 3 m and range 60 m due to the intersection of the surface bottom reflected path and the surface reflected path. Increasing the total number of paths by upgrading the instrumental set-up and adding more spatial diversity and more acoustic elements would help to increase the signal to noise ratio between effective target detection and background path intersections.

For each case, there is a good agreement between the simulated and the experimental images. Indeed, they suggest that the localization algorithm is robust to unexpected environment conditions often found in real coastal environments, potentially enhancing the future applicability of the method.

IV. CONCLUSION

This paper presents a simple, robust algorithm that addresses the problem of imaging quiet underwater intruders entering a harbor. The algorithm was tested using an experimental setup configured to represent an acoustic tripwire. The experimental apparatus was composed of a set of sources and receivers vertically spanning opposite sides of a harbor entrance or chokepoint.

The algorithm exploits the shadows produced at the receivers by one or more targets crossing the acoustic field. Each such shadow, which is expected to be strongest along specific acoustic paths and for specific source-receiver combinations, is continuously monitored. The algorithm is able to detect crossing targets and identify both the range and depth of the targets using a simple Gaussian–Ray-based construction. Theoretical computation shows that the diversity of the acoustic "view points" provides a complete and optimal acoustic net under varying environmental conditions. The multipath propagation is exploited to provide a dense acoustic net of higher quality than previously possible. The resultant image is a function of the number of acoustic elements (i.e., sources and receivers), the center frequency and band-

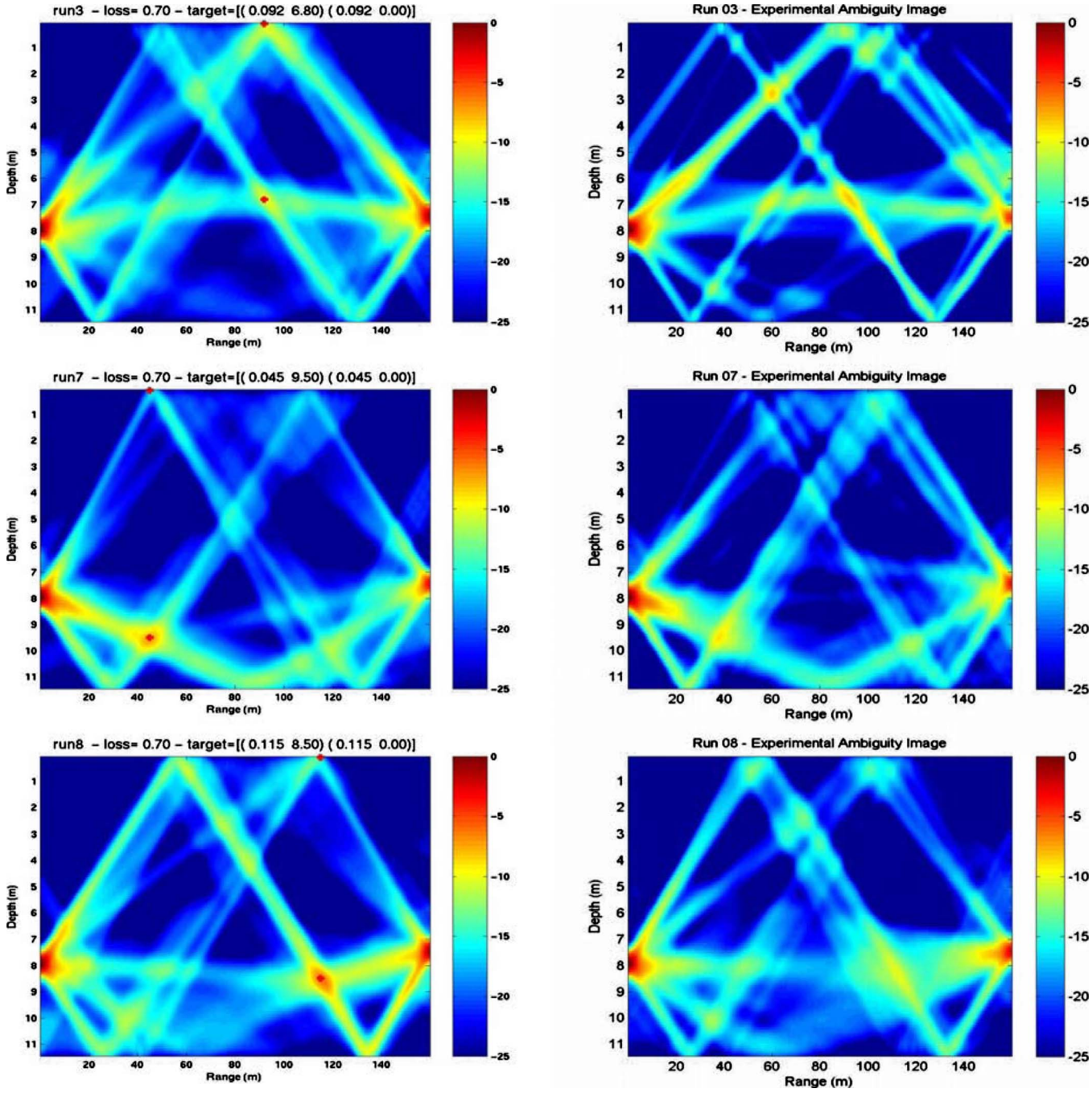


FIG. 6. (Color online) The 6 above plots represent experimental and theoretical ambiguity images (in dB) obtained for the BARRIER04 sea trial set-up. The left column shows the simulated images. The right column corresponds to the real data. Each pair (theoretical and real in a row) corresponds to a specific range and depth combination of the target: 6.8 m depth and 92 m range (top), 9.5 m depth and 45 m range (middle), 8.5 m depth and 115 m range (bottom). The underwater target is a 0.5 m diameter sphere, and the contrast coefficient is chosen $\gamma=4$. The theoretical image has been calculated including a 0.5 m diameter surface buoy located at the same range, from which the target was suspended.

width, and the number of acoustic paths. The concept is based upon forward propagation, hence, the level of the ambient noise is significantly less of an issue than might be expected using other techniques in noisy environment like harbors, or industrial areas. The only condition under which an object would not be detectable is if it were acoustically transparent (i.e., if it displayed neither absorption nor diffraction). Furthermore, the method is valid for multiple targets and therefore compatible with surface ship traffic. The choice of an optimal frequency is related to the size of the targets to be detected: the amplitude of the losses measured along each acoustic path depends on the interaction between the acoustic waves and the target itself. In the frame of harbor surveillance, there is no need to reach the very high frequencies used by backscattered systems, since the resolution does not directly depend on the acoustic frequency, but rather on the geometry of the acoustic web.

The algorithm was tested using acoustic data gathered at sea during the BARRIER-04 experiment. The experimental set-up did not provide an ideal data set to test the algorithm, especially in terms of number-of-sensors resolution. Nevertheless, the agreement between expected theoretical perfor-

mance and results validated both the concept and the method. The algorithm was validated for surface and submerged targets crossing at several depths and ranges.

The acoustic field created between the set of sources and receivers has a given width, a variable whose relative importance still needs to be assessed in future work. It is reasonable to expect that the time spent by any intruder to cross the barrier might give a fair estimation of its speed, useful information for tentative classification of the target. Potentially, air bubbles produced by a diver would stay even longer in the field, and the proposed method might allow tracking their displacement toward the surface. Future continuation of the work may be to apply the algorithm to a well-defined geometry in a harbor context. This barrier concept allows the control of a limited area (vertical plane) with an optimized level of confidence due to its robustness to environmental changes. However, the multiplication of acoustic barriers in a single harbor extends the control to a greater area, eventually up to the full monitoring of a three-dimensional area. It is envisioned that this system could also be used as an adjunct to other types of existing surveillance systems based, for instance, on backscattering. In this context, such systems might act as a first level of alarm to control a broader area—basically a volume outside the harbor where the noise level is significantly reduced, but with less accuracy and robustness due to their sensitivity to environmental changes. This concept would then be adopted for permanent installation, controlling critical access to certain parts of a harbor.

ACKNOWLEDGMENTS

The authors are grateful to the many NURC individuals who contributed to the successful conduct of the experiment, including Pierangolo Boni, Alberto Figoli, Marco Paoli, Bruno Miaschi, Claudio Olivero, Rosanna Giglione, and Roberto Lombardi. The experiments were conducted with the help of the crew of the NATO Research Vessel Leonardo. Alessandra Tesei and Mario Zampolli offered particularly helpful insight on forward target scattering.

[1]H. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, and P. Guerrini, "Demonstration of a high frequency acoustic barrier with a time-reversal mirror," IEEE J. Ocean. Eng. **28**(2), 246–249 (2003).
[2]A. L. Matveev, V. V. Mityugov, and A. I. Potapov, "Space-time acoustic holography of moving inhomogeneties," Acoust. Phys. **47**, 202–207 (2001).
[3]V. A. Zverev, A. L. Matveev, and V. V. Mityugov, "Matched filtering of acoustic diffraction signal for incoherent accumulation of a vertical antenna," Acoust. Phys. **41**, 518–521 (1995).
[4]S. M. Gorskii, V. A. Zverev, A. L. Matveev, and V. V. Mityugov, "Incoherent accumulation of acoustic diffraction signal," Acoust. Phys. **41**, 190–197 (1995).
[5]V. A. Zverev, P. I. Korotin, A. L. Matveev, V. V. Mityugov, D. A. Orlov, B. M. Salin, and V. I. Turchin, "Experimental studies of sound diffraction by moving inhomogeneities under shallow water conditions," Acoust. Phys. **47**, 184–193 (2001).
[6]W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).
[7]A. Parvulescu, "Matched-signal ('MESS') processing by the ocean," J. Acoust. Soc. Am. **98**, 943–960 (1995).
[8]W. A. Kuperman, T. Akal, W. S. Hodgkiss, S. Kim, G. Edelmann, and H. C. Song, "Forward scatter barrier with a time-reversal mirror," J. Acoust. Soc. Am. **108**, 2607 (2000).
[9]A. Tesei, H. C. Song, P. Guerrini, P. Roux, W. S. Hodgkiss, T. Akal, M. Stevenson, and W. A. Kuperman, "A high-frequency active underwater acoustic barrier experiment using a time reversal mirror; model-data comparison," in *High Frequency Ocean Acoustics*, edited by M. B. Porter, M. Siderius, and W. A. Kuperman (American Institute of Physics, Melville, NY, 2004), pp. 539–546.
[10]C. F. Gaumond, D. M. Fromm, J. F. Lingevitch, R. Menis, G. F. Edelmann, D. C. Calvo, and E. Kim, "Demonstration at sea of the decomposition-of-the-time-reversal-operator technique," J. Acoust. Soc. Am. **119**, 976–990 (2006).
[11]L. Carin, H. Liu, T. Yoder, L. Couchman, B. Houston, and J. Bucaro, "Wideband time-reversal imaging of an elastic target in an acoustic waveguide," J. Acoust. Soc. Am. **115**, 259–268 (2004).
[12]T. C. Yang, "Temporal resolutions of time-reversal and passive phase-conjugation for underwater acoustic communications," IEEE J. Ocean. Eng. **28**, 229–245 (2003).
[13]D. H. Chambers, "Analysis of the time-reversal operator for scatterers of finite size," J. Acoust. Soc. Am. **112**, 411–419 (2002).
[14]D. H. Chambers and J. G. Berryman, "Time-reversal analysis for scatterer characterization," Phys. Rev. Lett. **92**, 023902 (2004).
[15]T. Folegot, C. Prada, and M. Fink, "Resolution enhancement and separation of reverberation from target echo with the time reversal operator decomposition," J. Acoust. Soc. Am. **113**, 3155–3160 (2003).
[16]T. Folegot, D. Clorennec, J.-G. Minonzo, J. de Rosny, C. Prada, M. Fink, S. Hibral, and L. Berniere, "A high frequency time reversal array deployment in a very shallow water environment," J. Acoust. Soc. Am. **117**, 2462 (2005).
[17]C. Prada, J. de Rosny, D. Clorennec, J.-G. Minonzio, A. Aubry, M. Fink, L. Berniere, P. Billand, S. Hibral, and T. Folegot, "Experimental detection and focusing in shallow water by decomposition of the time reversal operator," J. Acoust. Soc. Am. **122**, 761–768 (2007).
[18]J. M. Stevenson, A. Tesei, P. Guerrini, P. Boni, P. Roux, H. C. Song, W. S. Hodgkiss, W. A. Kuperman, and T. Akal, "An underwater acoustic detection experiment based on forward scattering in a time reversal mirror," J. Acoust. Soc. Am. **116**, 2648 (2004).
[19]S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in side-scan sonar," IEEE J. Ocean. Eng. **28**, 90–105 (2003).
[20]J. Groen, R. E. Hansen, V. Myers, H. J. Callow, and M. A. Pinto, "Measurements on shadow contour estimation accuracy in synthetic aperture sonar images," Underwater Acoustics Measurements Conference, Crete, 2007.
[21]D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice Hall Signal Processing Series (Prentice Hall, Englewood Cliffs, NJ, 1993), Chap. 4.
[22]H. Weinberg and R. E. Keenan, "Gaussian ray bundles for modeling high-frequency propagation loss under shallow-water conditions," NUWC-NPT Technical Report 10,568, Newport, Rhode Island, 15 April 1996.

Analytical time-domain Green's functions for power-law media

James F. Kelly[a] and Robert J. McGough[b]
Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan 48824

Mark M. Meerschaert[c]
Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48824

(Received 29 April 2008; revised 19 July 2008; accepted 27 July 2008)

Frequency-dependent loss and dispersion are typically modeled with a power-law attenuation coefficient, where the power-law exponent ranges from 0 to 2. To facilitate analytical solution, a fractional partial differential equation is derived that exactly describes power-law attenuation and the Szabo wave equation ["Time domain wave-equations for lossy media obeying a frequency power-law," J. Acoust. Soc. Am. **96**, 491–500 (1994)] is an approximation to this equation. This paper derives analytical time-domain Green's functions in power-law media for exponents in this range. To construct solutions, stable law probability distributions are utilized. For exponents equal to 0, 1/3, 1/2, 2/3, 3/2, and 2, the Green's function is expressed in terms of Dirac delta, exponential, Airy, hypergeometric, and Gaussian functions. For exponents strictly less than 1, the Green's functions are expressed as Fox functions and are causal. For exponents greater than or equal than 1, the Green's functions are expressed as Fox and Wright functions and are noncausal. However, numerical computations demonstrate that for observation points only one wavelength from the radiating source, the Green's function is effectively causal for power-law exponents greater than or equal to 1. The analytical time-domain Green's function is numerically verified against the material impulse response function, and the results demonstrate excellent agreement.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977669]

PACS number(s): 43.35.Bf, 43.20.Hq, 43.20.Bi, 43.35.Mr [TDM] Pages: 2861–2872

I. INTRODUCTION

The attenuation coefficient for biological tissue may be approximated by a power law[1] over a wide range of frequencies. Measured attenuation coefficients of soft tissue typically have linear or greater than linear dependence on frequency. For example, breast fat has a power-law exponent of $y=1.5$, while breast tissue ranges[1] from $y=1$ to $y=1.5$. This nonquadratic frequency-dependent loss conflicts with the classical thermoviscous attenuation and, by the Kramers–Kronig relations, predicts a frequency-dependent phase speed or dispersion. Mathematically, the power-law frequency dependence of the attenuation coefficient cannot be modeled with standard dissipative partial differential equations with integer-order derivatives such as the Stokes wave equation[2] or the telegrapher's equation.[3] However, fractional partial differential equations (FPDEs) sucessfully capture this power-law frequency dependence.

Power-law attenuation has been incorporated in the time-domain via FPDEs, which add loss to the wave equation with a time-fractional derivative[4–7] or a space-fractional derivative.[8,9] Unlike loss models that utilize integer-order derivatives,[10,11] these FPDEs support power-law attenuation coefficients for a range of noninteger exponents y. The Szabo wave equation is one such FPDE that interpolates between the telegrapher's equation ($y=0$) and the Blackstock equation ($y=2$), allowing a large range of lossy behavior to be encapsulated in an expression with two loss parameters: an attenuation coefficient and a power-law exponent.[4] Frequency-domain formulations, based on the Kramer–Kronig relations[12–14] between attenuation and dispersion, have also been proposed for linear frequency dependence[15] and power-law dependence.[16–18]

The bulk of the above work is numerical, which motivates additional analytical modeling of the combined effects of power-law loss and diffraction with Green's functions in the time-domain. Unfortunately, few closed-form solutions that describe the effects of dispersion in the time domain have been reported. To date, approximate closed-form Green's function solutions have only been provided for the special cases of $y=0$ (frequency-independent media) and $y=2$ (viscous media). In order to construct Green's functions for general power-law exponents y, stable law probability distributions, which have previously been used to study fractional diffusion, are required. These stable distributions facilitate analytical descriptions of power-law attenuation directly in the time domain.

The purpose of this paper is to derive analytical time-domain three dimensional (3D) Green's functions for power-law media for $0<y<2$ in terms of stable probability densities. These Green's functions are exact solutions for power-law media and approximate Green's functions to the Szabo wave equation. In Sec. II, the Szabo wave equation and the power-law wave equation corresponding to these solutions are formulated as FPDEs. 3D Green's functions are then calculated in Sec. III, allowing solutions to be obtained for frac-

[a]Electronic mail: kellyja8@msu.edu
[b]Electronic mail: mcgough@egr.msu.edu
[c]Electronic mail: mcubed@stt.msu.edu

tional power-law media in terms of the Fox H-function[19] and the Wright function.[20] These functions have previously been used in the study of fractional diffusion,[21] fractional relaxation,[22] and fractional advection dispersion.[9] In the special cases of $y=0$, $1/3$, $1/2$, $2/3$, $3/2$, and 2, Green's function is expressed in terms of standard functions (Dirac delta, exponential, Airy, hypergeometric, and Gaussian functions). In the remaining cases, asymptotic expressions are provided. Numerical results and discussions of these results are provided in Secs. IV and V, respectively, followed by the conclusion in Sec. VI and the Appendix.

II. FPDE FORMULATIONS OF THE SZABO AND POWER-LAW WAVE EQUATIONS

A. Szabo wave equation

The Szabo wave equation[4] approximates power-law media with an attenuation coefficient given by

$$\alpha(\omega) = \alpha_0 |\omega|^y. \tag{1}$$

The Szabo wave equation was originally derived as an integro-differential equation for fractional power-law media. Subsequently, Ref. 23 expressed the Szabo wave equation in terms of fractional derivatives, thereby allowing the machinery of fractional calculus to be utilized. For $y \neq 1$, the Szabo wave equation is written as

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} - \frac{2\alpha_0}{c_0 \cos(\pi y/2)}\frac{\partial^{y+1} p}{\partial t^{y+1}} = 0, \tag{2}$$

where the third term accounts for dispersive loss. In the special cases of $y=0$ and $y=2$, the fractional derivative term reduces to first and third temporal derivatives, respectively. For fractional y, the third term is defined by the Riemann–Liouville fractional derivative defined in Appendix A.

For fractional y, Eq. (2) is a FPDE that approximates lossy dispersive media satisfying a power-law. As discussed in Ref. 4, Eq. (2) is valid for $\alpha_0 \ll 1$ in the (a) high-frequency limit for $y<1$ and (b) low-frequency limit for $y>1$. As shown below, a more general model for power-law media is obtained by including a higher-order temporal derivative. As in the case of the Stokes wave equation, the particle velocity and velocity potential also satisfy Eq. (2). From the definition of the fractional derivative given in Appendix A, Eq. (2) is also recognized as a singular integro-differential equation, where the fractional derivative term depends on the past history of the pressure $p(\mathbf{r},t)$.

B. Frequency-independent ($y=0$) and viscous ($y=2$) media

In the special cases of frequency-independent media ($y=0$) and viscous media ($y=2$), the Szabo wave equation reduces to well-known integer-order PDEs which are solved via standard methods. The $y=0$ case corresponds to the telegrapher's equation

$$\nabla^2 p - \frac{1}{c_\infty^2}\frac{\partial^2 p}{\partial t^2} - \frac{2\alpha_0}{c_\infty}\frac{\partial p}{\partial t} = 0, \tag{3}$$

which models one-dimensional (1D) damped string motion and electromagnetic wave propagation in conductive media. The reference frequency c_∞ is the phase speed in the limit of infinitely high frequency, and α_0 is the attenuation coefficient for $\omega=1$. The 3D Green's function for Eq. (3) consists of two terms: an exponential attenuated spherical wave, which is localized at time R/c_∞, and a wake involving a modified Bessel function, which extends for infinite time. If α_0 is small, the modified Bessel function term is of order α_0^2, yielding the high-frequency approximation[3]

$$g(R,t) \approx u(t) e^{-\alpha_0 R}\frac{\delta(t-R/c_\infty)}{4\pi R} \tag{4}$$

where $u(t)$ is the Heaviside step function. Equation (4) is an attenuated spherical wave, which is localized in time and hence is nondispersive.

In the viscous case ($y=2$), the Szabo equation reduces to the Blackstock equation

$$\nabla^2 p - \frac{1}{c_z^2}\frac{\partial^2 p}{\partial t^2} + \frac{2\alpha_0}{c_z}\frac{\partial^3 p}{\partial t^3} = 0, \tag{5}$$

which models acoustic wave propagation in viscous media[24] under a plane wave approximation for small α_0 and for small frequencies. The reference frequency c_z is the phase speed in the limit of zero frequency. Since the spatial bandwidth of a wavefield is large near the source, Eq. (5) fails to capture the high-frequency content in the extreme near field of a radiator.[2,24] The 3D Green's function of Eq. (5) is calculated via standard Fourier transform techniques, yielding

$$g(R,t) \approx \frac{1}{4\pi R}\sqrt{\frac{1}{4\pi R\alpha_0}}\exp\left(-\frac{(t-R/c_z)^2}{4R\alpha_0}\right), \tag{6}$$

which is an approximate solution to Eq. (5). Similar to the asymptotic solution to the Stokes wave equation,[25] Eq. (6) predicts Gaussian spreading of a spherical wave as the field radiates away from the source. However, Eq. (6) is noncausal, which means that Eq. (6) specifies nonzero field values for $t<0$. However, for α_0 small and R sufficiently large, Eq. (6) and the causal Green's function for the Stokes wave equation [see Eq. (3) in Ref. 26] are virtually indistinguishable, which is demonstrated numerically in Ref. 2 for the 1D case.

By utilizing the loss operator defined in Eq. (A2), the Szabo wave equation interpolates between the telegrapher's equation and the Blackstock equation. The telegrapher's equation is causal and hyperbolic, whereas the Blackstock equation is noncausal and parabolic. Mathematically, Eq. (2) is a second-order hyperbolic equation for $y<1$. For $y>1$, however, Eq. (2) becomes an order $y+1$ parabolic equation where high-frequency components propagate with infinite velocity. As $y \to 1^-$ in Eq. (2), $c_\infty \to \infty$, and as $y \to 1^+$ in Eq. (2), $c_z \to \infty$, which agrees with the transition between hyperbolic and parabolic behaviors at $y=1$. Thus, the Szabo wave equation interpolates between the two different behaviors as the power-law exponent y varies from 0 to 2.

C. Power-law wave equation

As discussed in Ref. 4, Eq. (2) is an approximate model for wave propagation in a power-law medium for small values of α_0. To facilitate the analytical solution of wave propagation in power-law media, a FPDE that exactly describes power-law attenuation for arbitrary α_0 is derived, and the Szabo wave equation is an approximation to this FPDE. This exact FPDE is then solved for a point source in space time, thereby yielding desired Green's function solution. Since both temporal and spatial Fourier transforms are utilized extensively throughout the derivation, the Fourier transform given by Eq. 3 in Ref. 4 is utilized throughout this paper. In order to derive the analytical time-domain Green's function solution, a power-law dispersion relationship relating wavenumber k and angular frequency ω is required. This dispersion relationship should yield (1) a power-law attenuation coefficient and (2) a frequency-dependent phase speed. The power-law dispersion relationship that satisfies these requirements is

$$k = \frac{\omega}{c_0} - \frac{\alpha_0(-i)^{y+1}\omega^y}{\cos(\pi y/2)}, \tag{7}$$

for $\omega \geqslant 0$ and $k(-\omega)=k^*(\omega)$ to ensure real solutions. The imaginary part of Eq. (7) yields the power-law attenuation coefficient given by Eq. (1). The real part produces

$$\frac{1}{c(\omega)} = \frac{\mathrm{Re}\ k(\omega)}{\omega} = \frac{1}{c_0} + \alpha_0 \tan\left(\frac{\pi y}{2}\right)|\omega|^{y-1}. \tag{8}$$

Letting c_1 denote the phase speed at $\omega_0=1$ yields the expression

$$\frac{1}{c(\omega)} = \frac{1}{c_1} + \alpha_0 \tan\left(\frac{\pi y}{2}\right)(|\omega|^{y-1} - 1), \tag{9}$$

which is valid for all $y \neq 1$. Equation (9) corresponds to the phase velocities computed via the Kramers–Kronig relations[14,27] and the time-causal theory.[13] In addition, Eq. (8) is in close agreement with the experimental dispersion data presented in Refs. 12–14 and 17. Thus, Eq. (2) supports the power-law attenuation and dispersion that is predicted by the Kramers–Kronig relationships and supported by experimental measurements.

In order to invert the wavenumber-frequency relationship into space time, the fourfold Fourier transform $\hat{P}(\mathbf{k},\omega)$ of the time-domain pressure $p(\mathbf{r},t)$ is multiplied by the dispersion relationship. Squaring both sides of Eq. (7) and multiplying by $\hat{P}(\mathbf{k},\omega)$ yield

$$\left[-k^2 + \frac{\omega^2}{c_0^2} - \frac{2\alpha_0}{c_0\cos(\pi y/2)}(-i\omega)^{y+1} - \frac{\alpha_0^2}{\cos^2(\pi y/2)}(-i\omega)^{2y}\right]\hat{P}(\mathbf{k},\omega) = 0. \tag{10}$$

Performing an inverse spatial Fourier transform produces the power-law Helmholtz equation

$$\nabla^2\hat{p} + \left[\frac{\omega^2}{c_0^2} - \frac{2\alpha_0}{c_0\cos(\pi y/2)}(-i\omega)^{y+1} - \frac{\alpha_0^2}{\cos^2(\pi y/2)}(-i\omega)^{2y}\right]\hat{p} = 0. \tag{11}$$

Finally, performing an inverse temporal Fourier transform using Eq. (A3) yields the power-law FPDE

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} - \frac{2\alpha_0}{c_0\cos(\pi y/2)}\frac{\partial^{y+1}p}{\partial t^{y+1}} - \frac{\alpha_0^2}{\cos^2(\pi y/2)}\frac{\partial^{2y}p}{\partial t^{2y}} = 0, \tag{12}$$

which satisfies the dispersion relationship, Eq. (7), exactly. For small α_0, the fourth term in Eq. (12) is negligible, thus yielding the Szabo wave equation given by Eq. (2).

III. 3D GREEN'S FUNCTION

Time-domain 3D Green's functions for power-law media are derived in this section. These Green's functions are exact solutions to the power-law wave equation in Eq. (12) and approximate solutions to the Szabo wave equation in Eq. (2). Applying a point source at time $t=0$ to Eq. (12) and considering Green's function $g(R,t)$ yield

$$\nabla^2 g - \frac{1}{c_0^2}\frac{\partial^2 g}{\partial t^2} - \frac{2\alpha_0}{c_0\cos(\pi y/2)}\frac{\partial^{y+1}g}{\partial t^{y+1}} - \frac{\alpha_0^2}{\cos^2(\pi y/2)}\frac{\partial^{2y}g}{\partial t^{2y}} = -\delta(\mathbf{R})\delta(t), \tag{13}$$

where $\mathbf{R}$ is the relative displacement between the source and the observer and $R=|\mathbf{R}|$. Applying a temporal Fourier transform to Eq. (13) yields

$$\nabla^2\hat{g} + k^2(\omega)\hat{g} = -\delta(\mathbf{R}), \tag{14}$$

where $k(\omega)$ is given by Eq. (7). The well-known Green's function for Eq. (14) is given by

$$\hat{g}(R,\omega) = \frac{e^{ik(\omega)R}}{4\pi R}. \tag{15}$$

Inserting Eq. (7) into Eq. (15) and grouping terms yield

$$\hat{g}(R,\omega) = \left[\frac{\exp(i\omega R/c_0)}{4\pi R}\right][\exp(-\alpha_0 R(|\omega|^y - i\tan(\pi y/2)\omega|\omega|^{y-1}))], \tag{16}$$

where the first factor solves the lossless Helmholtz equation. The time-domain Green's function for $y \neq 1$ is recovered by applying an inverse Fourier transform and the convolution theorem, yielding

$$g(R,t) = \mathcal{F}^{-1}[\hat{g}(R,\omega)] = \left[\frac{\delta(t-R/c_0)}{4\pi R}\right]\mathcal{F}^{-1}[\exp(-\alpha_0 R(|\omega|^y - i\tan(\pi y/2)\omega|\omega|^{y-1}))], \tag{17}$$

which is expressed as a temporal convolution

$$g(R,t) = g_D(R,t) * g_L(R,t). \tag{18}$$

Within the framework of linear time-invariant filters, Eq. (18) combines the effects of dispersive loss and diffraction as a cascade of two system functions. The first factor in Eq. (18) is the Green's function for the lossless wave equation given by

$$g_D(R,t) = \frac{\delta(t - R/c_0)}{4\pi R}, \tag{19}$$

which is responsible for the effect of propagation and diffraction. The second term is a loss function defined via

$$g_L(R,t) = \mathcal{F}^{-1}[\exp(-\alpha_0 R(|\omega|^y - i\tan(\pi y/2)\omega|\omega|^{y-1}))]. \tag{20}$$

Interestingly, the inverse Fourier transform defined by Eq. (20) may be calculated in closed form using the machinery of stable distributions.[28] The expression within the square brackets in Eq. (20) is identified as the characteristic function of a stable distribution with index y, center 0, skewness 1, and scale $(R\alpha_0)^{1/y}$. The inverse transform given by Eq. (20) is then expressed concisely as

$$g_L(R,t) = \frac{1}{(\alpha_0 R)^{1/y}} \tilde{f}_y\left(\frac{t}{(\alpha_0 R)^{1/y}}\right), \tag{21}$$

where the function $\tilde{f}_y(t)$, which is independent of R and α_0, is the maximally skewed stable distribution of index y with scale 1. This family of functions has been studied extensively.[29–31] Software is available for the computation of stable probability density functions (PDFs) and cumulative distribution functions[32] based on formulas developed in Ref. 33. In the following analysis, an alternate parametrization $f_y(t)$ is defined via

$$f_y(t) = |\sec(\pi y/2)|^{1/y} \tilde{f}_y(|\sec(\pi y/2)|^{1/y} t), \tag{22}$$

in order to utilize existing expressions for the stable density $f_y(t)$. In the case of both sublinear power-law media ($y<1$) and superlinear power-law media ($y>1$), $f_y(t)$ possesses the following analytical properties: (1) $f_y(t) \geqslant 0$ for all t, (2) $f_y(t)$ is infinitely continuously differentiable (or smooth), (3) $f_y(t)$ is unimodal, and (4) $\int_{-\infty}^{\infty} f_y(t)dt = 1$. Although these properties are common to all maximally skewed stable distributions, the behavior of $f_y(t)$ differs for $y<1$ and $y>1$. Properties of these functions are studied in the following two subsections for sublinear ($y<1$) and superlinear ($y>1$) power-law media. The special case of $y=1$ is discussed in Sec. III D.

Equation (18) is *approximate* Green's function, valid for small α_0, for the Szabo wave equation given by Eq. (2). However, Eq. (18) is an *exact* solution to the power-law FPDE given by Eq. (12) for linear homogeneous media since the imaginary part of Eq. (7) exactly describes the power-law behavior that the Szabo wave equation approximates. Thus, analytical Green's function given by Eq. (18) is not restricted to small α_0 and large R.

A. Sublinear power-law media ($0<y<1$)

For the general case of $0<y<1$, $f_y(t)$ was first considered in Ref. 34 and later in Ref. 35 as an inverse Laplace transform. Rewriting Eq. (22) as a inverse Laplace transform yields

$$f_y(t) = \mathcal{L}^{-1}\{e^{-s^y}\} = \frac{1}{2\pi i}\int_L e^{-s^y} e^{st} ds, \tag{23}$$

where L is a contour in the left half plane. The notation is Eq. (23) was previously utilized in Ref. 36. The following analytic properties of $f_y(t)$ for $0<y<1$ are known:[29,30] (1) $f_y(t)$ has support on $[0,\infty)$ and (2) as $t\to\infty$, $f_y(t) \sim t^{-y-1}$. The inverse Laplace transform defined by Eq. (23) is expressed in terms of the Fox H-function[19] using formulas presented in Ref. 22. The Fox H-function, which is defined in terms of a contour integral by Eq. (B1), generalizes many of the special functions, such as Bessel, hypergeometric, and Mittag-Leffler.[20] Using Eq. (62) in Ref. 22, the fractional exponential possesses the representation

$$\exp(-s^y) = \frac{1}{y} H_{0,1}^{1,0}\left(s \middle| \begin{matrix}(1,1)\\(0,1/y)\end{matrix}\right), \tag{24}$$

and the inverse Laplace transform is then calculated using Eq. (30) in Ref. 22, yielding

$$f_y(t) = \frac{1}{yt} H_{1,1}^{1,0}\left(\frac{1}{t} \middle| \begin{matrix}(0,1)\\(0,1/y)\end{matrix}\right). \tag{25}$$

For $y=0$, $1/3$, $1/2$, and $2/3$, Eq. (25) possesses the following exact representations:

$$f_0(t) = \delta(t), \tag{26a}$$

$$f_{1/3}(t) = \frac{u(t)}{(3t^4)^{1/3}} \mathrm{Ai}\left[\frac{1}{(3t)^{1/3}}\right], \tag{26b}$$

$$f_{1/2}(t) = u(t)\frac{1}{2\sqrt{\pi t^{3/2}}} \exp\left(-\frac{1}{4t}\right), \tag{26c}$$

$$f_{2/3}(t) = u(t)\frac{2}{3\sqrt{\pi t^2}} \exp\left(-\frac{4}{27t^2}\right) {}_2F_0\left(\frac{1}{6}, -\frac{1}{6}; ; -\frac{27t^2}{4}\right), \tag{26d}$$

where $\mathrm{Ai}(z)$ is the Airy function and ${}_2F_0(a,b;;z)$ is the confluent hypergeometric function of the second kind.

For the general case, an asymptotic expression, valid for $t \ll 1$, is derived using the asymptotic form[19] of $H_{1,1}^{1,0}(z)$. Alternatively, the inverse Laplace transform may be approximated via the method of steepest descents,[36] yielding

$$f_y(t) \approx u(t)\frac{A}{t^\nu} \exp\left(-\frac{B}{t^\mu}\right). \tag{27}$$

The coefficients A, B, ν, and λ depend only on y and are given by

$$A = \frac{y^{1/(2-2y)}}{\sqrt{2\pi(1-y)}}, \tag{28a}$$

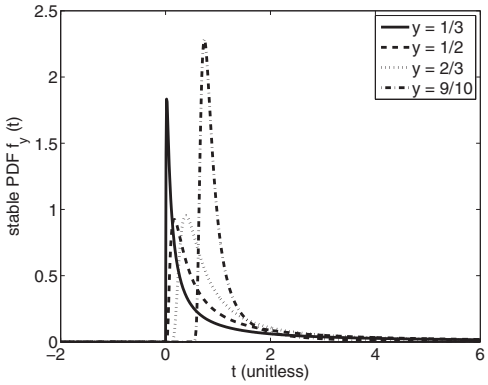


FIG. 1. Plots of maximally skewed stable PDFs $f_y(t)$ for four sublinear attenuation cases: $y=1/3$, $1/2$, $2/3$, and $9/10$. Equations (26a)–(26d) are evaluated for $y=1/3$, $1/2$, and $2/3$, while the $y=9/10$ case is evaluated with the STABLE toolbox (Ref. 32).

$$B = y^{1/(1-y)}\frac{1-y}{y}, \tag{28b}$$

$$\nu = \frac{2-y}{2-2y}, \tag{28c}$$

$$\mu = \frac{y}{1-y}. \tag{28d}$$

Equations (18), (27), and (28a)–(28d) provide a closed-form asymptotic approximation to the time-domain 3D Green's function for sublinear power-law media ($y<1$). These equations exactly satisfy Eq. (26c) at all times t for $y=1/2$ and are valid for $y\neq 1/2$ when $t\ll 1$. For $y\approx 1/2$, Eq. (27) also provides an excellent approximation[37] for all values of t. For $t>0$, f_y is infinitely continuously differentiable (or smooth); moreover, $f_y^{(n)}(t)=0$ as $t\to 0^+$ for all $n\geq 0$, implying strong causality.[2,36] However, for $y\neq 1/2$ and $t\gg 1$, the asymptotic solution given by Eq. (27) decays exponentially for large times, while $f_y(t)$ decays algebraically. Therefore, for large t, the first term in the asymptotic series for Eq. (25) is utilized, yielding[29]

$$\tilde{f}_y(t) \approx \frac{\sin(\pi y/2)\Gamma(y+1)}{\pi t^{y+1}}, \tag{29}$$

which indicates a slowly decaying tail.

To show the behavior of $f_y(t)$ for $0<y<1$, the PDFs for the sublinear attenuation cases $y=1/3$, $1/2$, $2/3$, and $9/10$ are displayed in Fig. 1. Since the ordinate $f_y(t)$ is a PDF, the abscissa t is unitless. The expressions given by Eqs. (26a)–(26d) are evaluated for $y=1/3$, $1/2$, and $2/3$, while the STABLE toolbox is utilized for $y=9/10$. In all of these cases, $f_y(t)$ is identically zero for $t<0$ and smooth. As y increases from $1/3$ to $9/10$, the main plume spreads out, while the asymptotic decay accelerates from $t^{-4/3}$ to $t^{-19/10}$ due to the behavior predicted by Eq. (29).

As $y\to 0$, $f_y(t)\to\delta(t)$, while as $y\to 1^-$, $f_y(t)\to\delta(t-1)$. Thus, there is a delay incorporated into the stable distribution which accounts for the slower phase velocity in a dispersive medium. This delay is apparent in the asymptotic representation given by Eq. (27). For very small t, the argument of the exponent in Eq. (27) is very large, causing $f_y(t)$ to approach zero. The delay t_d of $f_y(t)$ can be approximated by setting the exponential argument of Eq. (27) equal to unity, yielding $t_d\approx B^{1/\mu}$. Letting y vary from zero to 1, t_d is also seen to increase smoothly from zero to one.

B. Superlinear power-law media ($1<y\leq 2$)

In superlinear power-law media, the qualitative behavior of the loss function $g_L(R,t)$ differs from that observed in sublinear power-law media. The maximally skewed stable distributions possess the additional analytical properties for $1<y\leq 2$: (1) $f_y(t)$ has support on $(-\infty,\infty)$, (2) as $t\to\infty$, $f_y(t)\sim t^{-y-1}$ for $y<2$, and (3) as $t\to-\infty$, $f_y(t)$ decays with exponential order. For $y>1$, $f_y(t)$ is nonzero for *all* times t. Since the 3D Green's function is simply a delayed and scaled version of $f_y(t)$, the power-law Green's function $g(R,t)$ is nonzero for all t and is therefore noncausal for superlinear power-law media for all values of y such that $1<y\leq 2$. Since $f_y(t)$ decays with exponential order as $t\to-\infty$, the solution is very small. This property, which was discussed by Szabo[4] for the quadratic case ($y=2$), where $f_y(t)$ is Gaussian, is also true for all $1<y\leq 2$ according to the stable law properties listed above.

The stable distribution $f_y(t)$ is evaluated in terms of the Fox H-function [see Eq. (2.13) in Ref. 38] via

$$f_y(t) = \frac{1}{y}H_{1,1}^{1,0}\left(-t\left|\begin{matrix}(1-1/y,1/y)\\(0,1)\end{matrix}\right.\right), \tag{30}$$

where $1<y\leq 2$. Equation (30) may be simplified in terms of the Wright function $\phi(a,b;z)$, yielding

$$f_y(t) = \frac{1}{y}\phi\left(-\frac{1}{y},1-\frac{1}{y};t\right), \tag{31}$$

which is a known solution of the time-fractional diffusion problem.[39] For the special cases of $y=3/2$ and $y=2$, Eq. (31) is expressed in terms of the confluent hypergeometric function of the second kind and a Gaussian, respectively:

$$f_{3/2}(t) = \begin{cases} \dfrac{3}{4\sqrt{\pi}t^{5/2}}\,{}_2F_0\left(\dfrac{7}{6},\dfrac{5}{6};;-\dfrac{27}{4t^3}\right) & \text{if } t>0 \\ -\dfrac{2\exp(4t^3/27)}{3\sqrt{\pi t}}\,{}_2F_0\left(\dfrac{1}{6},-\dfrac{1}{6};;\dfrac{27}{4t^3}\right) & \text{if } t<0 \end{cases} \tag{32a}$$

$$f_2(t) = \frac{\exp(-t^2/4)}{\sqrt{4\pi}}. \tag{32b}$$

Note that the $y=2$ case in Eq. (32b) yields Eq. (6), which approximately solves the Blackstock equation[24] in Eq. (5).

Unlike the solution described earlier for sublinear power-law attenuation ($y<1$), $f_y(t)$ is nonzero for all t. Therefore, three separate cases are considered: early time ($t\ll R/c_0$), late time ($t\gg R/c_0$), and times near the wavefront. To obtain an early-time estimate, Eq. (30) is expanded in a Taylor series.[38] If $t\ll R/c_0$, then $t_r\to\infty$, allowing an asymptotic representation of Eq. (30) to be utilized:[38]

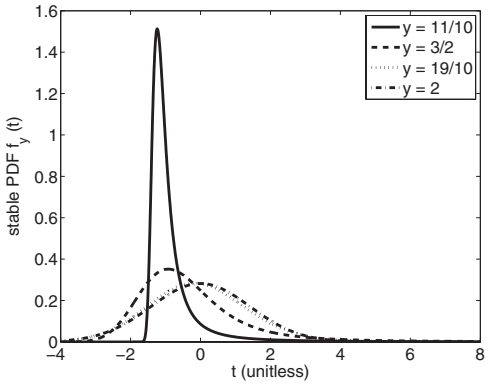


FIG. 2. Plots of maximally skewed stable PDFs $f_y(t)$ for four superlinear attenuation cases: $y=11/10$, $3/2$, $19/10$, and 2. Equations (32a) and (32b) are evaluated for $y=3/2$ and 2, while the $y=11/10$ and $19/10$ cases are evaluated with the STABLE toolbox (Ref. 32).

$$f_y(t) \approx A|t|^{\nu} \exp(-B|t|^{\mu}) \tag{33}$$

where

$$A = \frac{y^{-1/(2y-2)}}{\sqrt{2\pi(y-1)}}, \tag{34a}$$

$$B = y^{-y/(y-1)}(y-1), \tag{34b}$$

$$\nu = \frac{2-y}{2y-2}, \tag{34c}$$

$$\mu = \frac{y}{y-1}. \tag{34d}$$

For $y=2$, Eq. (33) reduces to $f_2(t)$ for all t, yielding the noncausal Blackstock solution. Equation (33) has the form of a "stretched exponential," which is a characteristic of anomalous diffusion.[40] In other words, the early-time response of superlinear power-law media is characterized by fractional diffusive behavior. For large t, the asymptotic formula given by Eq. (29) is utilized.

Since Eq. (33) is nonzero for all t, power-law Green's functions are noncausal for $1<y\leqslant 2$. However, the rapid exponential decay for $t \ll R/c_0$ makes the solution very small for large negative times ($t \ll -1$). Thus, the noncausal nature of Green's function for $1<y\leqslant 2$ is not evident in numerical evaluations for small α_0. From a physical perspective, the phase velocity becomes unbounded as frequency increases, implying that high-frequency components propagate infinitely fast. However, as noted in Ref. 2 for the special case of $y=2$, these high-frequency components experience high absorption and attenuate over a short distance, implying Green's function $g(R,t) \ll 1$ for $t \ll R/c_0$.

To show the behavior of $f_y(t)$ for $1<y\leqslant 2$, the stable PDFs for the superlinear power-law attenuation cases $y=11/10$, $3/2$, $19/10$, and 2 are displayed in Fig. 2. The expressions given in Eqs. (32a) and (32b) are evaluated for $y=3/2$ and 2, while the STABLE toolbox[32] is utilized for $y=11/10$ and $19/10$. In all cases, $f_y(t)$ is nonzero for all t and smooth. For $y=11/10$, $f_y(t)$ is skewed to the right and decays very rapidly for $t<-1$. As y increases from $11/10$ to 2, the PDF becomes increasingly symmetric, eventually converging to a Gaussian. Although $y=11/10$, $3/2$, and $19/10$ decay like t^{-y-1}, $y=2$ decays much more rapidly.

C. Linear power-law media ($y=1$)

The FPDE formulations of both the Szabo wave equation, given by Eq. (2), and the power-law wave equation, given by Eq. (12), break down as $y \to 1$ since $\cos(\pi y/2) \to 0$ in the denominator. To circumvent this problem, the frequency-domain solution at a reference frequency $\omega_0=1$ is considered. The phase velocity is computed by taking the limit of Eq. (9) as $y \to 1$. Without loss of generality, let y approach one from the right. Defining $y=1+\epsilon$ and letting $\epsilon \to 0$, then

$$\frac{1}{c(\omega)} = \frac{1}{c_1} + \alpha_0 \tan(\pi(1+\epsilon)/2)(|\omega|^{\epsilon}-1)$$
$$= \frac{1}{c_1} - \alpha_0 \frac{|\omega|^{\epsilon}-1}{\tan(\pi\epsilon/2)} \to \frac{1}{c_1} + \frac{2\alpha_0}{\pi\epsilon}(|\omega|^{\epsilon}-1), \tag{35}$$

where the limit $\tan z \to z$ is used in the last line. To finish the computation, the following limit is invoked:[41]

$$\frac{w^{\gamma}-1}{\gamma} \to \ln(w), \tag{36}$$

yielding

$$\frac{1}{c(\omega)} = \frac{1}{c_1} - \frac{2\alpha_0}{\pi}\ln|\omega|. \tag{37}$$

Equation (37) is identical to the Hilbert dispersive model derived in Ref. 15 and the Kramers–Kronig relation given by Eq. (13) in Ref. 14. Combining Eqs. (1) and (37) yields a dispersion relationship between the wavenumber k (spatial frequency) and the temporal frequency ω:

$$k(\omega) = i\alpha_0|\omega| + \frac{\omega}{c_1} - \frac{2}{\pi}\alpha_0\omega\ln|\omega|. \tag{38}$$

The 3D time-domain Green's function is computed by inserting Eq. (38) into the frequency-domain Green's function given by Eq. (15) and assigning the delay term ω/c_1 to $\hat{g}_D(R,\omega)$. Inverse Fourier transforming this product of diffraction and loss factors yields Eq. (18) with a loss function

$$g_L(R,t) = \mathcal{F}^{-1}\left\{\exp\left[-\alpha_0 R|\omega|\left(1+\frac{2i\,\mathrm{sgn}(\omega)\ln|\omega|}{\pi}\right)\right]\right\}, \tag{39}$$

where the signum function $\mathrm{sgn}(\omega)$ is introduced to enforce conjugate symmetry. Equation (39) is recognized as a stable distribution with index 1, center 0, skewness 1, and scale $\alpha_0 R$ using the parametrization in Ref. 28. Thus, the inverse Fourier transform is expressed as

$$g_L(R,t) = \frac{1}{\alpha_0 R}\tilde{f}_1\left(\frac{t}{\alpha_0 R}\right), \tag{40}$$

where $\tilde{f}_1(t)$ is the inverse Fourier transform of Eq. (39) with $\alpha_0 R=1$. Carrying out the transform by integrating over the negative and positive frequencies and expressing the cosine in terms of complex exponentials yield

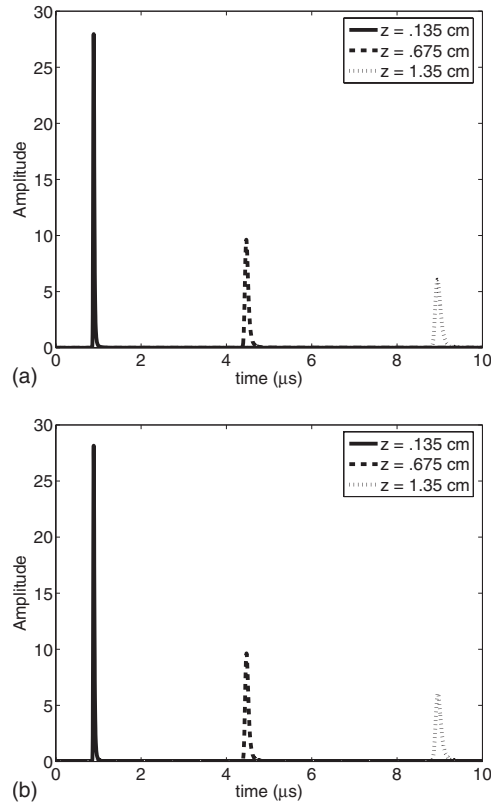


FIG. 3. Comparison of 3D Green's functions using (a) the MIRF approach and (b) the analytical Green's function approach given by Eq. (18). Green's functions are displayed at three different depths with $y=1.5$, $c_0=0.15$ cm/μs, and $\alpha_0=0.1151$ Np/MHz$^{1.5}$/cm.

$$\tilde{f}_1(t)=\frac{1}{\pi}\int_0^\infty \exp(-\omega)\cos\left[\omega t+\frac{2}{\pi}\omega\ln\omega\right]d\omega. \tag{41}$$

Although $\tilde{f}_1(t)$ cannot be expressed in terms of the Fox or Wright functions, this expression is computable using the STABLE toolbox.[32] Moreover, $\tilde{f}_1(t)$ has the same analytical properties as the superlinear power-law solutions shown earlier, including smoothness and exponential decay as $t\to-\infty$. Since $f_y(t)$ is nonzero for all t, the power-law Green's function is noncausal for $y=1$. For $t\gg1$, the asymptotic behavior of $\tilde{f}_1(t)$ is given by Eq. (29), which estimates the tail of the Green's function.

IV. NUMERICAL RESULTS

To numerically verify the analytical results presented above, the analytical Green's function was compared to the material impulse response function (MIRF) approach developed in Ref. 16. In Fig. 3(a), the 3D Green's function is computed in a power-law medium with parameters $y=1.5$, $c_0=0.15$ cm/μs, and $\alpha_0=0.1151$ Np/MHz$^{1.5}$/cm using the analytical formula in Eq. (18). In Fig. 3(b), the inverse Fourier transform defined by Eq. (20) was calculated numerically via an inverse fast Fourier transform (FFT). The two panels demonstrate excellent agreement. For power-law exponents $y=1/3$, $1/2$, $2/3$, $3/2$, and 2, the function $f_y(t)$ is represented in terms of the Airy, hypergeometric, exponential, and Gaussian functions. Equations (26a), (26b), (32a), and (32b) are evaluated numerically using the GNU Scientific Library.[42] For other values of y, the function $f_y(t)$ is represented as a maximally skewed stable distribution computed using the software package STABLE,[32] which is a general-purpose software for analyzing stable distributions. The results of STABLE have been independently verified using codes developed by Robinson[43] and a combination of asymptotic expressions, power series, and inverse power series. All demonstrate excellent numerical agreement. Transient fields are then evaluated using FFT-based convolutions.

Figure 4 shows the 3D power-law Green's function for $y=1/2$, $3/2$, and 2 for $\alpha_0=0.05$ mm^{-1} MHz^{-y}. The Green's functions are shown as functions of the radial coordinate R for times $t=20$, 30, 40, and 50 μs in order to emphasize the spatial distribution of acoustic energy in power-law media. The qualitative properties of sublinear power-law media ($y=0.5$), superlinear power-law media ($y=1.5$), and viscous media ($y=2$) are displayed in these plots. In these figures, the impulsive excitation has been stretched and smoothed by the filtering effect of the medium. The Green's function for sublinear power-law media ($y=0.5$) has a sharp wavefront at $R=c_0t$, where the field is identically zero to the right of this wavefront. Neither the superlinear power-law attenuation Green's function nor the viscous Green's function has this sharp wavefront. In the $y=1.5$ case, most of the energy is located in the slowly decaying tail of the Green's function, whereas in the viscous case ($y=2$), energy is symmetrically distributed about the location $R=c_0t$.

As time evolves in Figure 4, the peak amplitudes of all three of the Green's functions decrease, thus accounting for the transfer of acoustic energy into random thermal energy within the intervening medium. In the $y=0.5$ and $y=1.5$ cases, the duration of the Green's function increases, which is a consequence of the heavy tail behavior predicted by the asymptotic expression in Eq. (29). In the viscous case ($y=2$), spreading of the main plume occurs, although the Green's function is more localized relative to the fractional cases due to the rapid decay of the Green's function for large times. The absence of a slowly decaying tail is a consequence of the minimal dispersion associated with viscous and thermo-viscous loss.

To explore the effect of frequency-dependent attenuation on broadband pulse propagation, the velocity potential due to a point source at the origin was computed by convolving Eq. (18) with the pulse[44,45]

$$v(t)=A_0t^3\exp(-\beta t)\sin(2\pi f_0t)\mathrm{rect}\left(\frac{t}{W}\right). \tag{42}$$

The following simulations utilize the center frequency $f_0=2.5$ MHz, pulse length $W=1.2$ μs, damping factor $\beta=9.3750$ μs^{-1}, and a normalization factor of $A_0=616.3$ mm/μs. Two examples of velocity potential calculations are shown in two power-law media using parameters from Ref. 1: (1) a fat-like medium with $y=1.5$, $c_0=1.432$ mm/μs, and $\alpha_0=0.086$ Np/MHz$^{1.5}$/cm and (2) a liver-like medium with $y=1.139$, $c_0=1.569$ mm/μs, and $\alpha_0=0.0459$ Np/MHz$^{1.139}$/cm.

Figure 5 displays the velocity potential as a function of time t at radial distances 10, 25, 50, and 100 mm to show the

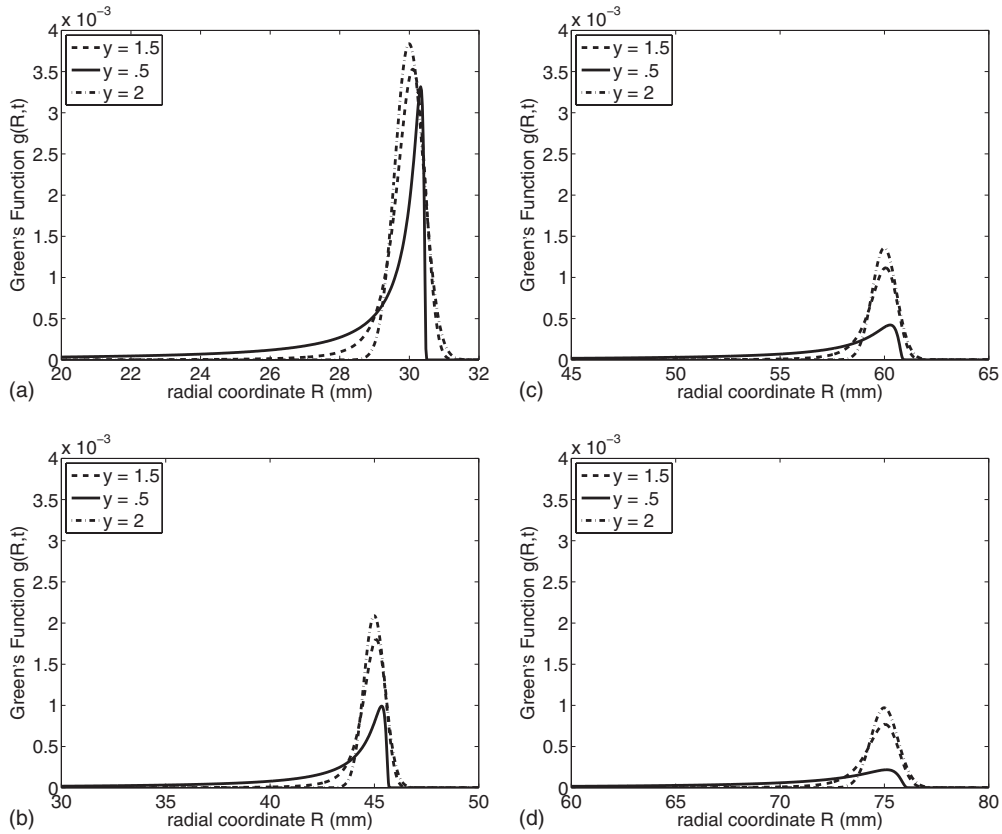


FIG. 4. Snapshots of the 3D power-law Green's function for $y=0.5$, 1.5, and 2.0 for $\alpha_0=0.05$ mm^{-1} MHz^{-y}. Snapshots of the Green's function are shown for $t=20$, 30, 40, and 50 μs.

effect of dispersion on the pulse shape. As R increases in Fig. 5, several qualitative features become evident. First, the pulse experiences rapid attenuation due to the combined effects of loss and spherical spreading. Second, the pulse changes shape due to the more rapid attenuation of the high-frequency components. In other words, the spectrum of the pulse experiences a frequency downshift. There are significant differences between the fat-like medium and the liver-like medium. The fat-like medium is more dissipative; therefore, greater attenuation and greater frequency downshifts are observed for large depths ($R=100$ mm). Although the liver-like medium is increasingly attenuated as the propagation distance increases, there is little change in the pulse shape. The pulse distortion in the fat-like medium indicates that the effects of dispersion become significant for large depths, which is consistent with the conclusion reached in Ref. 27.

V. DISCUSSION

A. Causality and the superlinear attenuation paradox

In Ref. 4, the "quadratic loss paradox" associated with the 1D solution of the Blackstock equation given by Eq. (5) is presented. In the discussion of Sec. VII of Ref. 4, Szabo notes that the 1D solution is noncausal yet provides an excellent approximation to the thermoviscous wave equation for small values of $\epsilon=\alpha_0 c_0|\omega|^{y-1}$, where ω is the upper frequency limit. Szabo's analysis is applied to three dimensions for superlinear power-law media ($y>1$).

By letting $\hat{R}=R/\lambda$, where λ is the wavelength associated with ω, the ratio of Green's function at time $t=0$ to the peak value is given by

$$\eta=\frac{f_y(-R/c_0/(\alpha_0 R)^{1/y})}{\max f_y(z)}=\frac{f_y(-\hat{R}^{1-1/y}\chi)}{\max f_y(z)} \tag{43}$$

where $\chi=|\sec(\pi y/2)|^{1/y}/(c_0\alpha_0^{1/y})$. Letting the normalized distance $\hat{R}=1$ in Eq. (43) yields the upper limit $20\log_{10}\eta<-136$ dB for $1<y\leqslant 2$. However, values of η are typically much smaller than this upper limit. For instance, for $y=1.5$, $20\log_{10}\eta<-360$ dB. Similar computations for other $y\geqslant 1$ demonstrate that for observation points only one wavelength from the radiating source, Eq. (18) is effectively causal for all $0<y\leqslant 2$.

B. Physical interpretation and application to ultrasonic imaging

Unlike the Green's function for lossless media, the Green's function for power-law media is not localized in space. Rather, the energy radiated by a point source is smeared out over a wide volume. For viscous media ($y=2$), the Green's function, given by a shifted and scaled Gaussian, decays rapidly for observation points distant from c_0R. For $y<2$, however, the Green's function decays as an inverse power of distance for observation points between the origin and R. This slow decay is interpreted as the gradual relaxation of the medium after the initial wavefront has passed. As time evolves from $t=20$ μs to $t=50$ μs in Fig. 4, the tail of

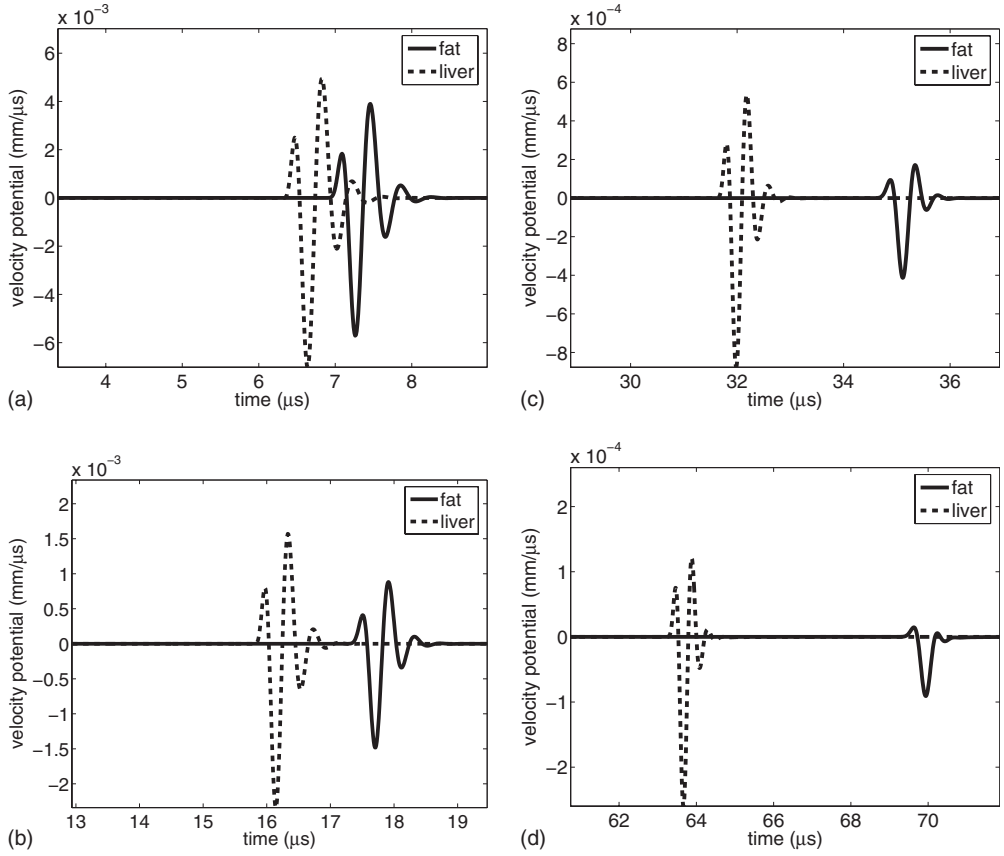


FIG. 5. Velocity potential produced by a point source excited by a broadband pulse defined by Eq. (42) in two power-law media: (1) a fat-like medium with $y=1.5$, $c_0=1.432$ mm/μs, and $\alpha_0=0.086$ Np/MHz$^{1.5}$/cm and (2) a liver-like medium with $y=1.139$, $c_0=1.569$ mm/μs, and $\alpha_0=0.0459$ Np/MHz$^{1.139}$/cm. The velocity potential is displayed at radial distances 10, 25, 50, and 100 mm.

the Green's function increases in both amplitude and duration, indicating that acoustic energy is being transferred from an organized spherical wavefront to a slowly decaying wake. Since this slow decay is not evident in the viscous case ($y=2$), a mechanism other than viscous dissipation (such as scattering by subwavelength structures) may be responsible for the observed wake.

Understanding the effect of attenuation on ultrasonic fields is important in B-mode image synthesis[46] and full-wave simulations in heterogeneous media.[47] Although the predicted change in phase velocity in most biological tissue is small,[12] the effects of dispersion accumulate with propagation distance and may alter pulse shape over acoustically large distances.[27] For instance, the change in phase velocity in the fat-like medium over the bandwidth of the pulse is less than 3 m/s, yet the incident pulse undergoes a significant frequency downshift at a depth of $R=100$ mm.

The 3D power-law Green's functions derived in Sec. III facilitate the study of 3D sound beams in dispersive power-law media. In particular, the impulse response produced by baffled circular and rectangular pistons may be constructed by integrating power-law Green's function over the radiating aperture. Specifically, the methodology developed in Ref. 26 for circular apertures in viscous media and Ref. 48 for rectangular apertures in viscous media may be extended to power-law dispersive media by utilizing the analytical Green's functions derived above. Since the power-law Green's function is efficiently evaluated using the STABLE toolbox,[32] the analytical expressions presented here may be used to either validate fields generated by dispersive full-wave linear solvers[7] or construct efficient single-scattering codes based on the Born approximation. Therefore, the analytical tools developed in this paper may further the understanding of the influence of dispersive loss on ultrasonic scattering and image formation.

Figure 5 provides an example of one such evaluation. For large depths ($R=50$ mm and $R=100$ mm), the distortion of the pulse is clearly evident in the fat-like medium but much less apparent in the liver-like medium. In fat, attenuation increases with frequency at a greater rate ($y=1.5$) than in liver ($y=1.139$), yielding a larger frequency downshift. Since attenuation and dispersion increase with frequency, increasing the center frequency f_0 of the pulse will increase both the absolute magnitude of frequency downshift and time delay, thereby degrading the ultrasound image in a diagnostic application. To address this problem, the analytical Green's functions presented in this paper may be used to select and evaluate pulse sequences for different imaging applications.

C. Comparison to frequency-domain models

To further validate the power-law Green's function, the Green's function for linear with frequency attenuation ($y=1$) was compared to the dispersive tissue model presented in Ref. 15, which incorporates a logarithmic frequency-

dependent phase delay that is equivalent to Eq. (37). Since Ref. 15 utilizes a 1D model, the $1/(4\pi R)$ factor in the power-law Green's function was omitted from this evaluation. The impulse response associated with the dispersive tissue model shown in Fig. 2 of Ref. 15 was then reproduced by evaluating Eq. (40) using the STABLE toolbox at times t retarded by a bulk delay given by $\tau_b = 1/c_1 = 6.67\ \mu s/cm$. The minimum phase τ_m presented in Ref. 15 is not required in the evaluation since this additional delay is incorporated into the stable density $\tilde{f}_1(t)$.

The analytical time-domain Green's functions presented in this paper may serve as references for calculations in a dispersive tissue model. These analytical models may prove advantageous relative to the frequency-domain models proposed in Refs. 15, 16, and 27, which require the evaluation of numerical inverse Fourier transforms. To expedite these inverse transforms, FFTs are utilized, which require (1) uniform discretization in time and (2) adequate sampling of $\hat{g}(R,\omega)$ in the temporal frequency domain. In contrast, the time-domain power-law Green's functions presented above are not limited by these two requirements. These Green's functions may be evaluated on a nonuniform grid of time samples without sacrificing efficiency. For example, these power-law Green's functions may be used to validate finite difference time domain codes which use nonuniform time stepping. In addition, the error tolerance associated with evaluating $f_y(t)$ via the STABLE toolbox is user specified,[32] thus making these analytical Green functions an ideal reference for MIRF and dispersive tissue model calculations.

D. Alternate fractional models

FPDEs have previously been utilized in hydrogeology to model advection due to microheterogeneity.[9,49] These FPDE models utilize a 1D advection-dispersion equation to capture the dispersion in heterogeneous aquifers. Like the solutions to the power-law wave equation, the solutions to the advection-dispersion equation are translated and scaled stable distributions, thus suggesting a connection between the hydrogeologic and ultrasonic models. Also, sound speed heterogeneity has been proposed as a mechanism for attenuation in marine sediments.[50] Since one possible mechanism for ultrasonic absorption is Class 0 scattering due to microheterogeneity,[51] the power-law wave equation may approximate scattering due to heterogeneity at the macromolecular level. The explicit connection between macromolecular scattering and power-law attenuation requires advanced mathematical tools such as homogenization theory and/or subordinated random processes.[30]

Alternate FPDE models for power-law attenuation have been reported, including models by Chen-Holm[8] and Wismer.[7] The Chen-Holm model utilizes a space-fractional derivative of order y, while the Wismer model utilizes a fractional time derivative of order $y-1$. Both the Chen-Holm and the Wismer models support power-law attenuation with exponent y in the zero-frequency limit, which is desirable for biomedical applications. Analysis of these models using the mathematical tools developed in this paper is an intended topic of future research.

VI. CONCLUSION

3D Green's functions in power-law media have been analytically computed in the time domain using the tools of fractional calculus. This paper shows that Green's functions for a medium with an attenuation coefficient given by Eq. (1) are maximally skewed stable distributions, shifted by a delay R/c_0, scaled by $(\alpha_0 R)^{1/y}$, and multiplied by a spherical diffraction factor of $1/(4\pi R)$. Mathematically, power law Green's functions are represented by a transient spherical wave, which embodies the diffractive process, convolved with a loss function $g_L(R,t)$, which embodies both the attenuation and dispersion of the medium. Physically, therefore, the solutions to the power-law wave equation are represented as the linear coupling of diffraction and dispersive loss. The time-domain power-law Green's functions presented here are exact solutions to the power-law wave equation in Eq. (12) and approximate solutions to the Szabo wave equation given by Eq. (2).

By solving for the time-domain power-law Green's function with the machinery of stable distributions, efficient numerical evaluation is available via the STABLE toolbox.[32] For all $y<1$, these Green's functions are expressed in terms of Fox H-functions, while for $y>1$, the Green's function is expressed in terms of the H function and Wright functions. For $y=0$, $1/3$, $1/2$, $2/3$, $3/2$, and 2, the Green's function solutions are expressed in terms of widely known special functions. The analytical Green's function was numerically verified against the MIRF result[16] and Hilbert dispersive model,[15] and example fields were computed. The results show that power-law Green's function is causal for $y<1$ and noncausal for $1 \leqslant y \leqslant 2$. Despite being noncausal for $1 \leqslant y \leqslant 2$, power-law Green's function provides an excellent causal approximation even for observation points very close to the radiating source. For $0<y<2$, the Green's function decays as t^{-y+1}, leaving a slowly decaying wake behind the primary wavefront. The Green's functions reported in this paper may be used to construct solutions to diffraction and scattering problems in power-law media.

ACKNOWLEDGMENTS

The authors thank John P. Nolan, Department of Mathematics and Statistics, American University and Stephen W. Wheatcraft, Department of Geological Sciences and Engineering, University of Nevada, Reno for useful discussions and advice. Geoff Robinson, CSIRO Mathematical and Information Sciences, is also acknowledged for making his maximally skewed stable distribution code available. This work was funded in part by NIH Grant No. 1R21CA121235 and NSF Grant No. DMS-0803360.

APPENDIX A: RIEMANN–LIOUVILLE FRACTIONAL DERIVATIVES

The Riemann–Liouville fractional derivative is formally defined via a hypersingular integral[52]

$$\frac{d^y f}{dt^y} = \frac{1}{\Gamma(-y)} \int_{-\infty}^{t} \frac{f(t')}{(t-t')^{1+y}} dt'. \tag{A1}$$

The nonintegrable singularity is removed from Eq. (A1) by recognizing a differentiation operator, yielding

$$\frac{d^{y+1} g}{dt^{y+1}} \equiv \begin{cases} \dfrac{1}{\Gamma(1-y)} \dfrac{d^2}{dt^2} \displaystyle\int_0^t \frac{g(t')}{(t-t')^y} dt' & \text{if } 0 < y < 1 \\ \dfrac{1}{\Gamma(2-y)} \dfrac{d^3}{dt^3} \displaystyle\int_{-\infty}^t \frac{g(t')}{(t-t')^{y-1}} dt' & \text{if } 1 < y < 2. \end{cases} \tag{A2}$$

In the literature, the $0<y<1$ case is known as the Riemann–Liouville form, whereas the $1<y<2$ case is the Liouville form. The following Fourier transform relationship for fractional derivatives is essential:[52]

$$\mathcal{F}\left(\frac{d^y g}{dt^y} \right) = (-i\omega)^y \hat{g}. \tag{A3}$$

In the $0<y<1$ case, Eq. (A3) assumes $g(t)$ vanishes for $t<0$.

APPENDIX B: THE FOX-*H* FUNCTION

The Fox H-function[20,52] is defined via a complex contour integral

$$H_{p,q}^{m,n}\left(z \left| \begin{matrix} (a_p, A_p) \\ (b_q, B_q) \end{matrix} \right. \right) \equiv \frac{1}{2\pi i} \int_L \chi(s) z^s ds, \tag{B1}$$

where $(a_p,A_p)=(a_1,A_1),(a_2,A_2),\ldots,(a_p,A_p)$, $(b_q,B_q)=(b_1,B_1),(b_2,B_2),\ldots,(b_q,B_q)$, and the integral density $\chi(s)$ is defined as

$$\chi(s) = \frac{\Pi_{j=1}^m \Gamma(b_j - B_j s) \Pi_{j=1}^n \Gamma(1 - a_j + A_j s)}{\Pi_{j=m+1}^q \Gamma(1 - b_j + B_j s) \Pi_{j=n+1}^p \Gamma(a_j - A_j s)}. \tag{B2}$$

Since the Gamma function is singular when the argument is a negative integer, Eq. (B2) contains an infinite number of poles in general. Therefore, the contour L is chosen to separate the poles of Eq. (B2). Note that Eq. (B1) is the inverse Mellin transform of Eq. (B2). The Fox H-function, which generalizes most of the special functions of mathematical physics, is discussed in great detail in Refs. 19 and 22.

APPENDIX C: THE WRIGHT FUNCTION

The Wright function $\phi(\alpha,\beta;z)$ is defined by an infinite series [see Eq. (1.11.1) in Ref. 52]:

$$\phi(\alpha,\beta;z) \equiv \sum_{k=0}^{\infty} \frac{1}{\Gamma(\alpha k + \beta)} \frac{z^k}{k!}. \tag{C1}$$

The Wright function and the more general H-function are related [see Eq. (6.2.66) in Ref. 52] via

$$H_{1,1}^{1,0}\left(z \left| \begin{matrix} (b,\alpha) \\ (0,1) \end{matrix} \right. \right) = \phi(-\alpha, b; -z). \tag{C2}$$

Equation (C2) is computed by evaluating the contour integral definition of the H-function, which possesses an infinite number of poles, with Cauchy's residue theorem. The resulting infinite series is identified as Eq. (C1).

[1] F. A. Duck, *Physical Properties of Tissue*, 1st ed. (Academic, London, 1990), pp. 99–124.
[2] M. J. Buckingham, "Causality, Stokes' wave equation, and acoustic pulse propagation in a viscous fluid," Phys. Rev. E **72**, 026610 (2005).
[3] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[4] T. L. Szabo, "Time-domain wave-equations for lossy media obeying a frequency power-law," J. Acoust. Soc. Am. **96**, 491–500 (1994).
[5] A. Hanyga and M. Seredyńska, "Power-law attenuation in acoustic and isotropic anelastic media," Geophys. J. Int. **155**, 830–838 (2003).
[6] M. Caputo, "Linear models of dissipation whose Q is almost frequency independent-II," Geophys. J. R. Astron. Soc. **13**, 529–539 (1967).
[7] M. G. Wismer, "Finite element analysis of broadband acoustic pulses through inhomogeneous media with power law attenuation," J. Acoust. Soc. Am. **120**, 3493–3502 (2006).
[8] W. Chen and S. Holm, "Fractional Laplacian time-space models for linear and nonlinear lossy media exhibiting arbitrary frequency power-law dependency," J. Acoust. Soc. Am. **115**, 1424–1430 (2004).
[9] R. Schumer, D. A. Benson, M. M. Meerschaert, and S. W. Wheatcraft, "Eulerian derivation of the fractional advection-dispersion equation," J. Contam. Hydrol. **48**, 69–88 (2001).
[10] S. Leeman, "Ultrasound pulse propagation in dispersive media," Ultrasound Med. Biol. **25**, 481–488 (1980).
[11] J. M. Blackledge and S. Leeman, "Green's functions for acoustic fields in dispersive media," J. Phys. D: Appl. Phys. **16**, L247–L250 (1983).
[12] M. O'Donnell, E. T. Jaynes, and J. G. Miller, "Kramers–Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701 (1981).
[13] T. L. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power-law," J. Acoust. Soc. Am. **97**, 14–24 (1995).
[14] K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, "On the applicability of Kramers–Krönig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).
[15] K. V. Gurumurthy and R. M. Arthur, "A dispersive model for the propagation of ultrasound in soft tissue," Ultrason. Imaging **4**, 355–377 (1982).
[16] T. L. Szabo, "The material impulse response for broadband pulses in lossy media," Proceedings of the IEEE Ultrasonics Symposium, Honolulu, HI 2003 (unpublished), pp. 741–751.
[17] P. He, "Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 114–125 (1998).
[18] N. V. Sushilov and R. S. Cobbold, "Frequency-domain wave equation and its time-domain solutions in attenuating media," J. Acoust. Soc. Am. **115**, 1431–1436 (2004).
[19] A. M. Mathai and R. K. Saxena, *The H-function with Applications in Statistics and Other Disciplines* (Wiley, New York, 1978).
[20] R. Gorenflo, Y. Luchko, and F. Mainardi, "Analytical properties and applications of the Wright function," Fractional Calculus Appl. Anal. **2**, 383414 (1999).
[21] R. Gorenflo and F. Mainardi, "Fractional calculus and stable probability distributions," Arch. Mech. **50**, 377–388 (1998).
[22] W. G. Glöckle and T. F. Nonnenmacher, "Fox function representation of non-Debye relaxation processes," J. Stat. Phys. **71**, 741–756 (1993).
[23] M. Liebler, S. Ginter, T. Dreyer, and R. E. Riedlinger, "Full wave modeling of therapeutic ultrasound: Efficient time-domain implementation of the frequency power-law attenuation," J. Acoust. Soc. Am. **116**, 2742–2750 (2004).
[24] D. T. Blackstock, "Transient solution for sound radiated into a viscous fluid," J. Acoust. Soc. Am. **41**, 1312–1319 (1967).
[25] D. G. Crighton, *Modern Methods in Analytical Acoustics: Lecture Notes* (Springer-Verlag, London, 1996).
[26] J. F. Kelly and R. J. McGough, "The causal impulse response for cicular sources in viscous media," J. Acoust. Soc. Am. **123**, 2107–2116 (2008).
[27] R. S. C. Cobbold, N. V. Sushilov, and A. C. Weathermon, "Transient propagation in media with classical or power-law loss," J. Acoust. Soc. Am. **116**, 3294–3303 (2004).
[28] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes* (Chapman and Hall, New York, 1994).
[29] J. P. Nolan, *Stable Distributions: Models for Heavy Tailed Data* (Birkhauser, New York, 2005).

[30]W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1966), pp. 165–173, and 548–549.
[31]V. M. Zolotarev, *One-Dimensional Stable Distributions* (American Mathematical Society, Providence, RI, 1986).
[32]J. P. Nolan, "Numerical calculation of stable densities and distribution functions," Commun. Stat. Stoch. Models **13**, 759–774 (1997); http://academic2.american.edu/~jpholan/stable/stable.html
[33]M. Kanter, "Stable densities under change of scale and total variation inequalities," Ann. Probab. **3**, 697–707 (1975).
[34]H. Pollard, "The representation of $e^{-z^{\lambda}}$ as a Laplace integral," Bull. Am. Math. Soc. **52**, 908–910 (1946).
[35]M. Dishon, J. T. Bendler, and G. H. Weiss, "Tables of the inverse Laplace transform of the function $e^{-s^{\beta}}$," J. Res. Natl. Bur. Stand., Sect. B **95**, 433–467 (1990).
[36]A. Hanyga and V. E. Rok, "Wave propagation in micro-heterogeneous porous media: A model based on an integral-differential wave equation," J. Acoust. Soc. Am. **107**, 2965–2972 (2000).
[37]A. Hanyga, "Propagation of pulses in viscoelastic media," Pure Appl. Geophys. **159**, 1749–1769 (2002).
[38]W. R. Schneider, "Stable distributions: Fox function representation and generalization," in *Stochastic Processes in Classical and Quantum Systems*, edited by S. Albeverio, G. Casati, and D. Merlini (Spring-Verlag, Berlin, 1985).
[39]F. Mainardi and G. Pagnini, "The Wright functions as solutions of the time-fractional diffusion equation," Appl. Math. Comput. **141**, 51–62 (2003).
[40]R. Metzler, W. G. Glöckle, and T. F. Nonnenmacher, "Fractional model equation for anomalous diffusion," Physica A **211**, 13–24 (1994).
[41]T. Kobayashi and S. Imai, "Spectral analysis using generalized cepstrum," IEEE Trans. Acoust., Speech, Signal Process. **32**, 1087–1089 (1984).
[42]M. Galassi, *GNU Scientific Library Reference Manual*, 2nd ed. (Network Theory Ltd., Bristol, UK, 2006).
[43]G. Robinson, "Rapid computations concerning log maximally-skewed stable distributions, with potential use for pricing options and evaluating portfolio risk," Private Report, 2005.
[44]N. Denisenko, G. Scarano, M. Matteucci, and M. Pappalardo, "An approximate solution of the transient acoustic field," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 821–827 (1985).
[45]J. L. San Emeterio and L. G. Ullate, "Diffraction impulse-response of rectangular transducers," J. Acoust. Soc. Am. **92**, 651–662 (1992).
[46]J. A. Jensen, D. Gandhi, and W. D. O'Brien, "Ultrasound fields in an attenuating medium," Proceedings of the IEEE Ultrasonics Symposium, Baltimore, MD, 1993, pp. 943–946.
[47]G. Pinton and G. Trahey, "Full-wave simulation of finite-amplitude ultrasound in heterogeneous media," Proceedings of the IEEE Ultrasonics Symposium, New York, 2007, pp. 130–133.
[48]J. F. Kelly and R. J. McGough, "Transient acoustic fields produced by rectangular apertures and linear arrays in viscous media," Proceedings of the IEEE Ultrasonics Symposium, New York, 2007, pp. 1553–1556.
[49]B. Baeumer, D. A. Benson, M. M. Meerschert, and S. W. Wheatcraft, "Subordinated advection-dispersion equation for contaminant transport," Water Resour. Res. **37**, 1543–1550 (2001).
[50]J. M. Collis, W. M. Carey, and A. D. Pierce, "A physical explanation for less than quadratic recorded attenuation values," OCEANS 2007 Europe, Aberdeen, Scotland, 18–21 June 2007, pp. 1–3.
[51]T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out* (Elsevier Academic, Burlington, MA, 2004).
[52]A. A. Kilbas, H. M. Srivastava, and J. J. Truhillo, *Theory and Applications of Fractional Differential Equations* (Elsevier, Amsterdam, 2006).

Scattering of the fundamental shear horizontal mode in a plate when incident at a through crack aligned in the propagation direction of the mode

M. Ratassepp[a)]
Department of Mechanics, Tallinn University of Technology, Ehitajate tee 5, Tallinn 19086, Estonia

M. J. S. Lowe[b)] and P. Cawley
Department of Mechanical Engineering, Imperial College, London SW7 2AZ, United Kingdom

A. Klauson
Department of Mechanics, Tallinn University of Technology, Ehitajate tee 5, Tallinn 19086, Estonia

(Received 17 March 2008; revised 13 August 2008; accepted 20 August 2008)

A study of the scattering of the fundamental guided wave SH_0 at a through-thickness narrow notch directed along the wave's propagation in a plate is presented. The results are obtained from Finite Element simulations and experimental measurements. Good agreement is found between the simulations and the measurements. The results are shown for a range of crack lengths and shapes. The scattered wave field consists of the reflected and diffracted SH_0 mode and also contributions from mode conversions to the S_0 mode. It is found that the coefficient of direct reflection of the SH_0 mode has an undulating nature depending on the length of the crack. This is caused by interference phenomena that are related to the interaction of different surface wave types generated on the crack surfaces and their diffractions at both tips of the crack. It is shown that the dominating part of this reflection is generated by the delayed "Rayleigh type" surface waves reflected from the far tip of the crack. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2987426]

PACS number(s): 43.35.Cg, 43.20.Mv, 43.35.Pt [YHB] Pages: 2873–2882

I. INTRODUCTION

The research is motivated by the desire to detect defects in large areas of structures. Among the various nondestructive testing (NDT) techniques, ultrasonic guided waves are becoming established for the inspection of large areas, for example, plate structures[1] and pipelines.[2] The advantage of using guided waves is that they can propagate along the plate structure to enable the monitoring of large areas from a single point. By generating waves in a chosen direction in the structure and then analyzing received reflections, it is then possible to locate and quantify any defects along the line of wave propagation. This reduces inspection time substantially and also enables the inspection of areas in structures that are not directly accessible (for example, insulated pipes). This is not possible with conventional local point-by-point scanning techniques.

The subject of this paper is to investigate the low frequency SH_0 mode interaction with an axial crack in a plate. In previous research, the SH_0 mode has been favored for practical NDT of plates.[3] Being nondispersive at all values of frequency and having the particle motion in the plane of the plate, this mode can travel long distances with negligible attenuation and without the signal shape changing. It is also unaffected by inviscid fluid contact with the structure. Furthermore, at lower frequencies it is the only mode in the shear wave family that can propagate in the plate, thus excluding multimode excitation by the transducers. The SH_0 mode is particularly attractive for the current investigation because the particle displacement is perpendicular to the plane of propagation and therefore it is expected to be sensitive to the cracks that are oriented along the propagation direction. Lamb modes such as A_0 and S_0 can also be used for the detection of cracks in plates[4,5] but they have the particle motion entirely in the plane of the propagation and are thus expected to be reflected less by axially aligned cracks. The SH_0 mode is also very similar in nature to the torsional mode T(0,1) in pipes, the only difference being the geometric curvature of the material. The T(0,1) mode is the main mode of choice for the established long range guided wave pipe inspection tools.[6] Therefore, this study can be helpful to practical pipe testing by improving understanding of the torsional mode interaction with axial cracks in pipes.

Studies of the interaction of the shear waves[7–12] and Lamb waves[4,5,13–15] with discontinuities such as thickness changes have already been performed by others. For example, Finite Element (FE) method simulations and experimental studies[4,5,13] have helped to understand the interaction of individual Lamb waves with a variety of defects. It was shown that the sensitivity of different wave modes to particular notches or cracks is dependent on the frequency-thickness product, the mode type, the mode order, and the geometry of the reflector. Also, analytical approaches such as modal decomposition have been useful to describe the role of nonpropagating modes on the reflection and transmission characteristics from thickness changes in planar waveguides using various wave modes.[10,11,15] A popular method has also

[a)]Electronic mail: madisr@staff.ttu.ee
[b)]Electronic mail: m.lowe@ic.ac.uk

been to combine FE and analytical techniques for analyzing the interaction of guided waves with arbitrary shaped planar cracks.[8,9] However, all these approaches have been focused on defects which are aligned normal to the direction of wave propagation and are therefore not suitable to explain the case of an axial crack which has a small, or indeed negligible, dimension in that direction.

Recent paper by Rajagopal and Lowe[12] showed that the analysis of the SH_0 mode scattering normally incident at a finite crack can be performed with a plane stress model using the FE method. Their study showed that at normal incidence to the face of the crack the reflected near field is strongly affected by the surface waves diffracting at the tips of the crack. In the present case, these surface waves play a much more important role because there is no straight reflection due to the limited or zero width of the notch. A parallel phenomenon can be seen from the classical problem where bulk shear waves interact with the boundary of a semi-infinite medium at grazing incidence.[16] Moreover, the problem closely resembles bulk shear wave scattering by internal cracks in an infinite media at low frequencies. In this case, the incident mode is converted to Rayleigh surface waves, which travel along the surface of the crack and are diffracted and reflected at the tips of the crack thus forming the wave field around the crack.[17]

The aim of this paper is to determine the scattering characteristics from an aligned through-thickness cracks in a plate when the low frequency SH_0 mode is incident. The results of this study will be essential to understand the physical nature of the phenomenon and provide some insight into the possibility of detecting and characterizing this kind of defect using ultrasonic guided waves. In this work, we use the FE and experimental approach of Rajagopal and Lowe but examine only the case when the guided wave source and the crack are aligned. The scattering is investigated for a range of crack lengths and shapes. The report focuses on the straight reflection back toward the source, which is monitored at several distances from the crack. Also, some results are shown for the diffraction of the SH_0 mode in the direction of 90° from the crack face. The predictions are validated by experiments. Finally, the influence of several surface waves on the scattering mechanism is investigated.

II. LOW FREQUENCY GUIDED WAVES IN A PLATE

The properties of Lamb waves and SH waves are well known.[18] Figure 1 presents the group velocity dispersion curves for a steel plate. These group velocity curves show the speed of propagation of a wave packet and are therefore useful for long range testing. They were calculated using the program DISPERSE.[19] In practice, the authors favor an approach to NDT in which the signals are kept as simple as possible and to avoid multimodality.[20] Therefore, the work presented here is limited to the very low frequency-thickness range below 400 kHz mm, where only three modes A_0, S_0, and SH_0 may exist. This relatively low frequency range is also of particular interest to the authors because this aids our understanding of the case of pipes with cracks where a low frequency torsional wave is used.[21]

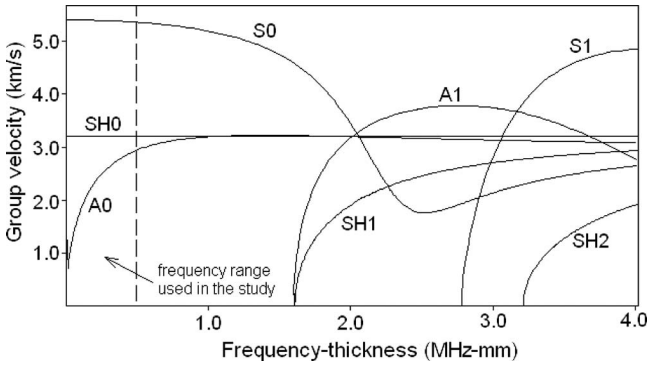


FIG. 1. Group velocity dispersion curves for shear and Lamb waves in a steel plate.

The SH_0 mode is very attractive for long range NDT. It is completely nondispersive at all frequencies and its phase velocity and group velocity are equal to the bulk shear wave velocity. The displacement and stress fields of this mode are also simple. In Fig. 2, the mode propagating in the x direction has only one displacement component, u_z, in the z direction and one component of stress. Also, its displacement and stress are uniform through the thickness of the plate so that its sensitivity to a defect is not dependent on the through-thickness location of the defect. The absence of normal displacement means that there is no attenuation if the plate is in contact with inviscid liquids. The main advantage of using the SH_0 mode in the current study is that its particle motion is perpendicular to the crack surfaces. Therefore, it might be sensitive to such defects and provide reflections that characterize the crack.

Although the S_0 and A_0 modes have been used efficiently in defect characterization,[4,5] they are not suitable for the detection of cracks aligned in the wave propagation direction. Both modes have particle motion which is entirely in the plane of the propagation and therefore obviously their wave field is much less affected by axially oriented defects. The S_0 mode, being the fundamental symmetric compressional wave, has a predominant in-plane motion in the direction of x. The A_0 mode represents a flexural type of mode having predominant out-of-plane displacement u_y; in-plane displacement u_x, and stresses σ_{xx} are antisymmetric with respect to the midplane of the plate. However, since these two waves may exist within the frequency range of interest, it is possible in principle for both of them to be excited by mode conversion when SH_0 is incident.

III. FINITE ELEMENT MODELING

The FE study was performed in a two-dimensional domain, with the assumption of plane stress, using the ABAQUS

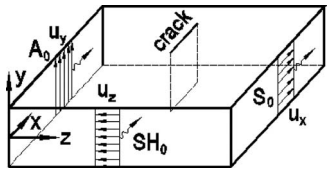


FIG. 2. Predominant displacement mode shapes of A_0, S_0, and SH_0 at low frequencies propagating in the x direction.

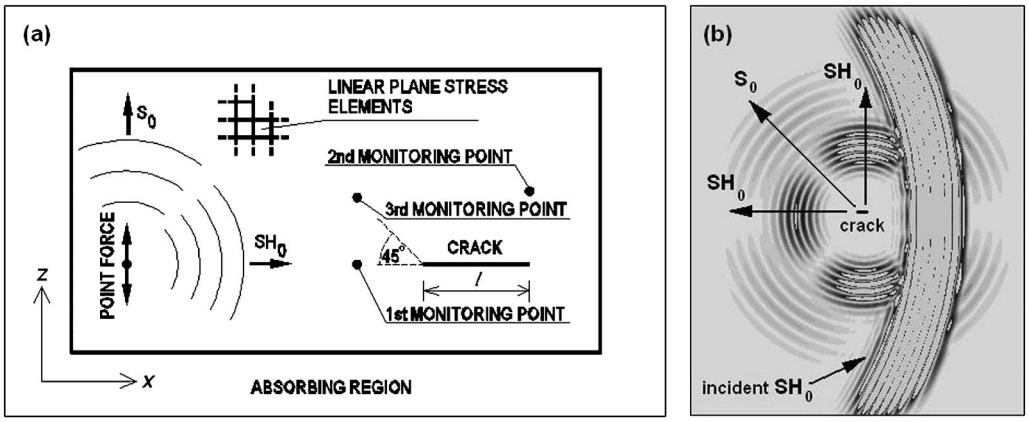


FIG. 3. (a) Setup of the FE model. (b) Typical FE time record showing the displacement amplitudes of reflected and diffracted waves around the crack after the interaction of the SH_0 mode with the 24 mm long crack.

software with its explicit time stepping procedure.[22] The plane stress approximation has been found to work very satisfactorily in earlier studies modeling the SH_0 mode propagation in plates.[12]

The spatial model of the problem is shown in Fig. 3(a). The properties of the material are shown in Table I. The mesh of the plate consists of perfectly square four noded linear elements with two degrees of freedom (displacements in x, z directions). More than 16 elements per wavelength of the SH_0 mode were used in the study, which satisfies the limit for accurate modeling.[12] Also, these elements describe satisfactorily the motion of the S_0 and SH_0 modes at low frequencies, since these modes have predominant in-plane displacement fields which are symmetric to the midplane of the plate and approximately uniform through the thickness. This model does not support the propagation of the bending type modes such as A_0. However, antisymmetric waves are not generated in the cases studied in this work because all of the features of the geometry and of the incident wave are symmetric. The plate is surrounded by an "absorbing region," which has recently been efficiently introduced into FE modeling.[23] The waves that enter this area are increasingly damped and eventually they die out. This helps to reduce model size and avoid unwanted reflections from the edges of the plate.

Variants of the spatial models were set up in order to model different cracks. Three geometries of the discontinuity were modeled.

(1) *Zero-width crack*. This crack type was created simply by disconnecting adjacent elements in the plate model. This approach to defining a crack has been used widely in previous FE studies of guided wave interactions with defects.[4,5,12]
(2) *Notch (rectangular crack)*. The crack is defined by removing elements from the mesh of the plate. The width of the elements along the crack line was also modified when necessary to obtain the desired width of the crack. Additional FE analyses with further refinement of the mesh (results not shown here) showed that the unequal element dimensions used in the simulations did not cause discrepancies in the wave propagation. However, care must be taken when modeling very narrow cracks. The reduced size of the element also increases the calculation time.
(3) *V crack*. This geometry is suitable for modeling very narrow cracks but with sharp tips.

TABLE I. Material properties for steel used in the FE model.

Density (ρ)	Poisson's ratio (ν)	Young modulus (E)
7.9 g/cm^3	0.3	210 GPa

Crack geometries (2) and (3) were important to compare the FE results with experiments where a finite width machined notch was used.

The low frequency excitation in the z direction was applied at the desired node on the crack axis. This generates circular waves of the SH_0 mode primarily propagating in the x direction and also the S_0 mode, which primarily propagates in the z direction. The displacement amplitude of both modes is strongest in their principal direction of propagation and decays according to a cosine angular pattern away from the main direction.[24] Although this means that the S_0 mode can be present in areas away from the crack axis, at low frequency thicknesses this mode travels much faster than SH_0 and can be time gated out. The point force excitation consisted of a 100 kHz, 5 cycle tone burst multiplied by a Hanning window. The propagation of the waves was simulated in the time domain. Explicit time integration and a fixed time step corresponding to stable integration were used for the simulation.[25]

The signals were monitored at three different positions [Fig. 3(a)]. The choice of these points was made after trial simulations and points were selected at locations where the scattered wave field was the strongest. A typical plot showing the contour of resultant displacements of different scattered waves after the interaction of the incident mode with the crack is illustrated in Fig. 3(b). In the first location, which is along the line of the crack and the excitation point, the displacement u_z of the reflected SH_0 mode was monitored. The second monitoring point is at 90° from that line, measured from the back tip of the crack, in order to monitor the displacement u_x of the diffracted SH_0 mode. In the third

position, which is at 45° to the incident direction and has the same x coordinate as the first monitoring point, the displacement component u_z of the reflected S_0 mode was monitored. The incident signal was always monitored at point 1. The model was repeated for different crack lengths. The main focus was directed toward scattering at cracks whose length was up to the wavelength of the incident mode, but the direct reflection was also investigated at long cracks up to 15 wavelengths in length.

Both modulus and phase of the reflected and diffracted signals were calculated. The modulus was obtained by dividing the amplitude of the reflected or diffracted signals by the amplitude of the incident signal in the frequency domain. To obtain the results for plane waves interacting with the crack, it is necessary to consider the beam spreading of the propagating waves. As the waves in the model decay cylindrically away from the source and also assuming the crack to act as the emitter of cylindrical waves, then the reduction of the amplitudes in both cases is inversely proportional to the square root of the propagation distance from the source. Thus,

$$\text{modulus} = \frac{A(f)\sqrt{r}}{I(f)\sqrt{d}}, \tag{1}$$

where $A(f)$ is the amplitude in the frequency domain of reflected or diffracted signals, $I(f)$ is the amplitude in the frequency domain of the incident signal, r is the distance between the tip of the crack and the monitoring point, and d is the distance from the source to the monitoring point. The phase was calculated only for the reflected and diffracted SH_0 signals and was obtained by using a simple fast Fourier transform algorithm.

IV. EXPERIMENTAL WORK

Experiments were conducted on steel plates of dimensions 1250 mm long, 1250 mm wide, and 1 mm thick. 0.7 mm wide through-thickness notches with lengths 4, 8, 12, and 70 mm were cut in the plate using a milling machine. The notch width was 15 times smaller than the wavelength of the incident wave for most of the frequencies used. Each crack was positioned approximately 800 mm from one end of the plate, with its axis lying at equal distances from the other sides of the plate, as shown in Figs. 4(a) and 4(b).

Excitation of the SH_0 mode was achieved using a wideband piezoelectric shear transducer (Panametrics V301, 0.5 MHz center frequency) coupled to the plate through a small area using a small brass disk (3 mm diameter, 0.4 mm thickness) and a shear coupling agent. The excited signal consisted of a 5 cycle tone burst in a Hanning window centered at the required frequency. Such narrow band signals were used in order to achieve good signal strength. For each crack length, the center frequency of excitation was varied from 200 to 400 kHz in order to obtain different values of length of the crack in terms of wavelength of the incident mode. This scaling is permissible because the SH_0 and S_0 modes in the low frequency range of study have velocity and mode shapes which are independent of the plate thickness; thus the only spatial dimensions of the problem are the crack

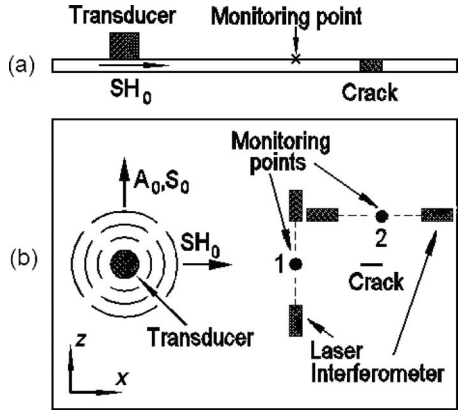


FIG. 4. Arrangement of the experimental test. (a) Side view; (b) plan view.

dimensions and the propagation distances. The signals were generated using a Wavemaker (Macro Design Ltd., UK) instrument.

The monitoring of the incident signal and detection of the reflected and diffracted signals was achieved using a laser interferometer (Polytec OFV 2700, with dual differential fiber optic lines) to measure the displacements at the surface of the plate at the required positions. Two laser beams were aligned at $+30°$ and $-30°$ to the normal, but focused on the same spot; thus the difference between their two signals gave the in-plane displacement. A thin reflective tape was attached to the surface of the plate to enhance the optical backscatter. The signals were in general quite weak and the quality of each measurement was enhanced by taking 500 averages and by applying a bandpass filter.

The excitation and detection positions were varied at each center frequency value, so that the distances between the crack and monitoring point and monitoring point and transducer remained a certain number of wavelengths. Typical experimental time records measured at the first and second monitoring points are shown in Figs. 5(a) and 5(b), respectively. The length of the crack was 12 mm, the excitation frequency was 200 kHz, the signal was generated at $22\lambda_{SH0}$ (SH_0 wavelengths at the center frequency of the tone burst) away from the crack, and the measurement distance from the

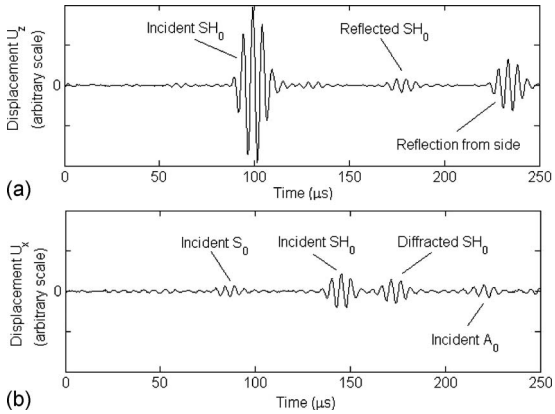


FIG. 5. Typical experimental signals; the center frequency is 200 kHz; the crack is 12 mm long. (a) Time record from the first monitoring point 118 mm away from crack, (b) time record from the second monitoring point 118 mm away from crack. Both signals are plotted to the same scale.

crack was $7\lambda_{SH0}$ for both measurement points. From the measurement in Fig. 5(a), it is seen that only the incident and reflected SH_0 modes are presented, but in Fig. 5(b) the measured signal at the second point contains also some S_0 and A_0 modes that were excited by the transducer. Differentiation of unwanted signals and the SH_0 mode may most easily be achieved by choosing sufficiently long propagation distances that their signals separate.

The modulus of the reflected and diffracted SH_0 modes for each crack length was calculated similarly as in the FE procedure by dividing the frequency spectrum of the reflected signal by that of the incident signal and compensating for beam spreading.

V. RESULTS AND DISCUSSION

A. Reflection and diffraction as a function of length of the crack

Using the results of the FE simulations, the scattering effect was studied with SH_0 mode incident. This was done over varying crack lengths in a plate. The excitation was located at a distance of $22\lambda_{SH0}$ and nodal displacements were monitored at $7\lambda_{SH0}$ from the crack. Figures 6(a)–6(c) show the typical FE time history records monitored at the first, second, and third point, as shown in Fig. 3(a). The defect is a 0.7 mm wide notch and its length is 24 mm ($\lambda_{SH0}/l = 0.75$, where l is the length of the crack). The figures show that the incident, reflected, and diffracted parts of the signals are extractable allowing time gating for the calculation of the modulus of reflection and diffraction.

The reflection and diffraction moduli were calculated for each crack length using Eq. (1) and the results of different excitation frequencies and crack lengths were normalized to the wavelength λ_{SH0} of the incident mode. In the FE study, the excitation frequency was 100 kHz and in the experiment this was varied from 200 to 400 kHz. Experimental measurements were taken at the appropriate excitation and monitoring distances for each frequency.

At first, we investigate the scattering at short cracks whose length is similar to the wavelength of the incident mode. This means that the direct reflection of SH_0 includes the reflection components at the tip and at the end of the crack which are not separable in the signal. Figure 7 shows the reflection modulus for monitored SH_0 modes at the first and S_0 mode at the third point. The predictions for direct SH_0 mode reflection were obtained for three crack cases: zero-width crack, 0.7 mm wide notch, and 0.7 mm wide V crack. As we see, the width of the crack slightly affects the reflection behavior. It is seen that the reflection from the zero-width crack is slightly weaker and its values are shifted toward higher crack length to wavelength ratios compared to other crack types used in the study. The simulations with a notch and with a V crack are similar and give the best agreement with the experimental results. The S_0 mode is clearly the weakest reflected signal and therefore it is not suitable for the crack characterization. However, it shows that the scattering phenomenon is accompanied by mode conversions. Due to symmetry, there is no reflection of the S_0 mode along the axis of the crack and the first monitoring point.

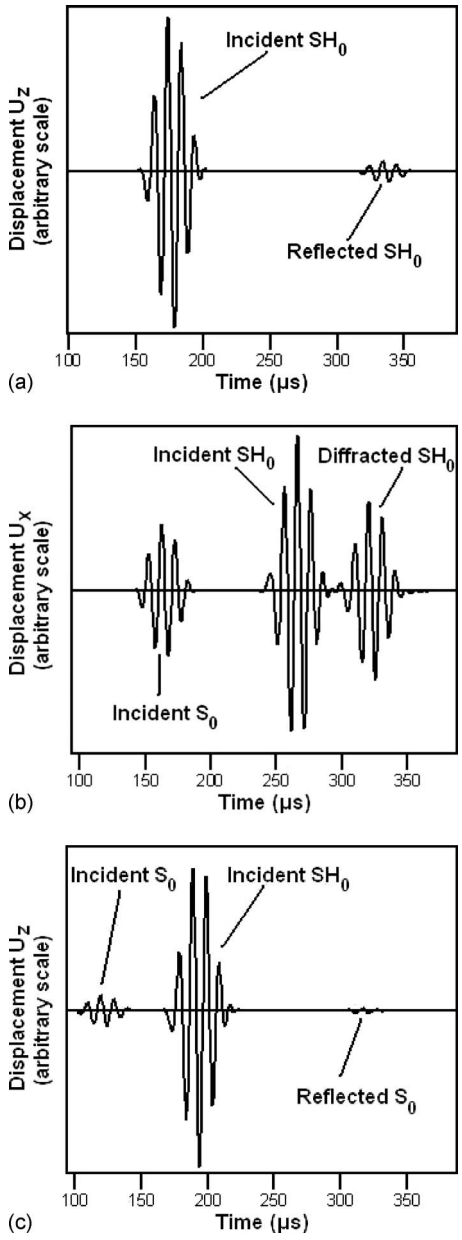


FIG. 6. Typical FE time records from the (a) first, (b) second, and (c) third monitoring points showing the reflected and diffracted signals; 0.7 mm wide notch 24 mm in length was used in this model; center frequency is 100 kHz.

The reflection modulus of the SH_0 mode at zero-width crack in Fig. 7 starts at a value of zero, and then rises as the crack length is increased. However, while we would expect the ratio to approach a constant value as the crack grows larger, there is an oscillation whose origin is not clear. As we can see, the first maximum reflection occurs when the length of the crack is about 45% of the incident wavelength λ_{SH0}, then there is a low point at around 63%, and then an increase again reaching a maximum at about 87%. Finally, the ratio starts to drop again. This behavior is due to the interference between the reflection from the near tip of the crack and the reflection slightly later from the far tip of the crack. This will be shown and discussed later. Their superposition in the resulting reflected wave packet may be constructive or destructive, depending on the duration of the delay. Thus, for example, the reflection from the far end of a crack of 45%

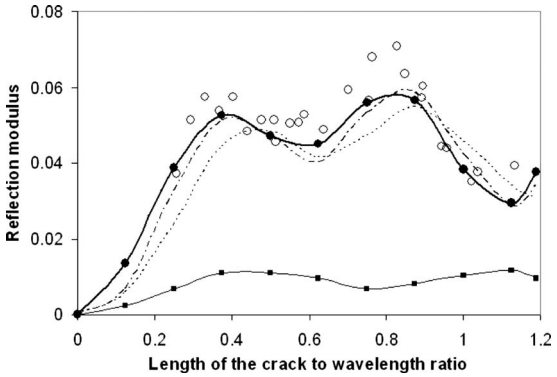


FIG. 7. Predicted variation of reflection modulus with length of the crack normalized to wavelength. Monitored SH_0 mode at the first point: FE results with a 0.7 mm width notch (solid circle), 0.7 mm width V crack (dash-dot), zero-width crack (dotted line), and experimental results (empty circle). Monitored S_0 at the third point—FE results (solid square).

wavelength arrives back with a delay corresponding to one wavelength, so that it is in the same phase as the reflection from the start of the crack and thus causes maxima. However, the maxima and minima do not occur precisely at multiples of 25% wavelength positions, but slightly smaller distances. The reason for this is that the SH_0 mode is partially converted to Rayleigh type surface waves, also known as pseudo-Rayleigh waves[26] on the stress free surfaces of the crack, which are reflected and diffracted at the crack tips. These waves are slower than the shear mode and have a smaller wavelength. Additionally, it is also worth noting that these interfering waves have different amplitudes. This explains a reflection minimum that is not zero at 63% of λ_{SH0}.

Figure 8 shows the diffraction modulus of monitored SH_0 modes at the second point, 90° from the crack face. The diffraction in this direction is stronger than the reflection in the incident direction. Here we can see a similar initial rise of the diffraction strength with the increase of the crack length. However, there is a maximum instead of a trough at 60% of λ_{SH0} compared with the reflection modulus of the direct reflection. Again Rayleigh type surface waves play an important role here. The diffraction is composed of two signals—one diffracted from the front tip of the crack and the other from the back tip of the crack. The second diffraction is delayed due to the surface wave propagation along the crack

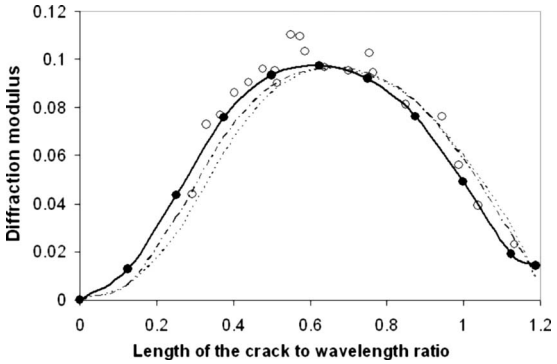


FIG. 8. Predicted variation of diffraction modulus with length of the crack normalized to wavelength. Monitored SH_0 mode at the third point: FE results with a 0.7 mm width notch (solid circle), 0.7 mm width V crack (dash-dot), zero-width crack (dotted line), and experimental results (empty circle).

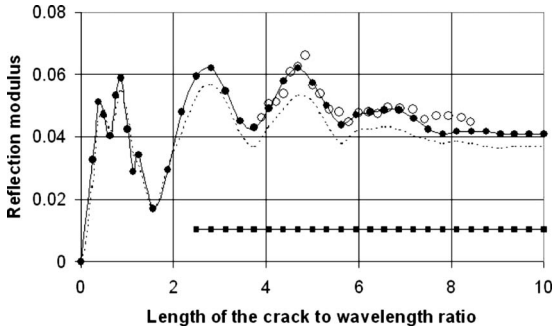


FIG. 9. Reflection modulus of SH_0 mode for short and long cracks. FE results: SH_0 near tip reflection at 0.7 mm V crack (solid square), SH_0 far tip reflection at 0.7 mm V crack (solid circle), and SH_0 far tip reflection at zero-width crack (dashed). Experimental SH_0 far tip reflection (empty circle).

faces. However, this time the diffraction is in phase with the first diffraction because there is no additional phase delay due to the absence of back propagation. Finally, the ratio starts to decrease. Again the finite width rectangular crack gives the best fit with the experimental results. However, the difference between the results with the different crack types is not significant.

Returning to Fig. 7 the reflection modulus ends with a minimum, which is much larger than the trough at 63% of λ_{SH0}. This behavior is not explained anymore by the observed interference pattern between the two reflected wave packets. To understand the reason for this, reflection characteristic from longer cracks were investigated. Figure 9 shows the reflection modulus for the SH_0 mode in the incident direction from cracks up to $10\lambda_{SH0}$ in length. Predictions were made for the zero-width crack and the V crack. There is a very good agreement between modeling and experimental results. Again we see that the reflection from the zero-width crack is slightly weaker compared to the other results.

Now we can easily separate the near tip reflection, which is notably smaller than the delayed far tip reflection. Its amplitude is independent of the crack length. Interestingly the modulus of the far tip reflection, even if it is separated from the near one, remains periodically undulating with weakening amplitude. The interval between peaks is longer for long cracks than for short cracks (in a wavelength range). The main reason for this is that the interference phenomenon is complicated considerably by the presence of another wave generated at the surfaces of the crack, whose reflection is not separable from the reflected SH_0 in short crack cases. However, the reflection of these waves can be seen separately in cases of long cracks, as shown in Figs. 10(a) and 10(b). Here it is important to note that the separation of reflected signals also depends on the duration of the tone burst. In the case of a longer tone burst, the crack would also have to be longer before it is possible to separate the reflections. These figures show the FE and experimental time records of SH_0 wave interaction with a $8.1\lambda_{SH0}$ long crack. The two reflections are clearly separable—the smaller reflection at the near tip is followed by the second stronger reflection from the far tip, induced by Rayleigh type surface waves. However, we can also see small amplitude waves arriving before the second reflection. These waves are generated due to the presence of

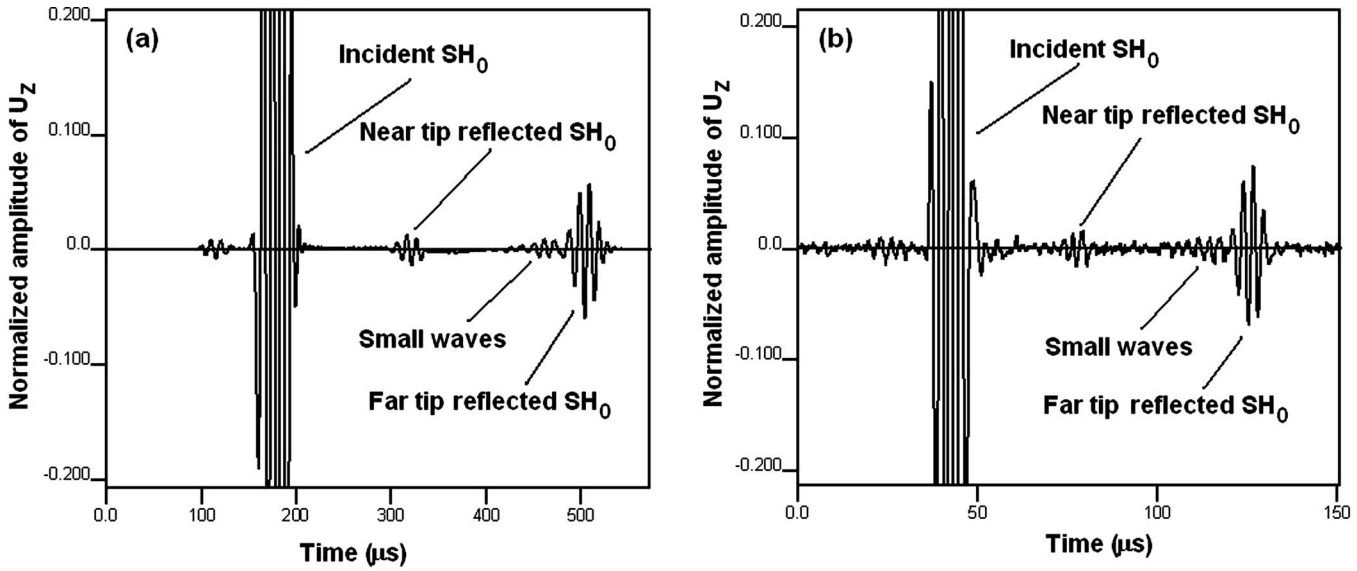


FIG. 10. Typical time domain signals showing the reflections at long cracks. (a) FE measurement at 262 mm ($8.1\lambda_{SH0}$) long crack, center frequency 100 kHz; (b) experimental measurement at 70 mm ($8.1\lambda_{SH0}$) long crack, center frequency 390 kHz.

"leaky type" surface waves on the crack, which also cause small reflection of the SH_0 mode. These waves are faster than Rayleigh type surface waves and are generated due to the S_0 mode interacting with the crack surfaces; this is a similar phenomenon to that of bulk pressure wave conversion to leaky surface waves on the free boundary of a semi-infinite medium.[27] Quantitative analyses will follow later. The weakening oscillation in Fig. 9 is explained by the fact that this wave is coupled into a plane shear wave that propagates away from the crack. Due to this, the surface wave loses energy and, thus, decays in its propagation direction. Therefore, the strength of the final diffraction weakens as the crack length is increased and this also decreases the amplitude of undulations in superimposed signals.

B. Influence of the distance of source and measurement location on direct reflection

In order to see how the distance between the source, receiver, and the crack influences the reflection of the SH_0 mode, FE predictions were made for a couple of source and receiver points along the crack axis. The distance between the source and measurement point was chosen to be $5\lambda_{SH0}$, while the distance between the source and the crack was varied through four values: 10, 20, 30, and $40\lambda_{SH0}$. The crack length was again varied using a number of simulations and the reflection modulus was calculated according to Eq. (1). Figure 11 shows that the reflection modulus for all crack length to wavelength ratios decreases with the increasing source distance from the zero-width crack. The reason for this is that the reflected waves show simple cylindrical decay away from the crack. It is clearly seen that the reflection modulus for different source and measurement locations has the same variance when the crack length to wavelength ratio changes. The placement of peaks and troughs is similar for all curves. Moreover, the ratio of modulus M_1/M_2 of any two arbitrarily chosen source location results 1 and 2, at each crack length, can be expressed as the square root inverse ratio of source distances from the crack $\sqrt{(r_2+d_2)/(r_1+d_1)}$. This assumption is valid when the measurements have been made in the far field. In Fig. 12, the reflection modulus for

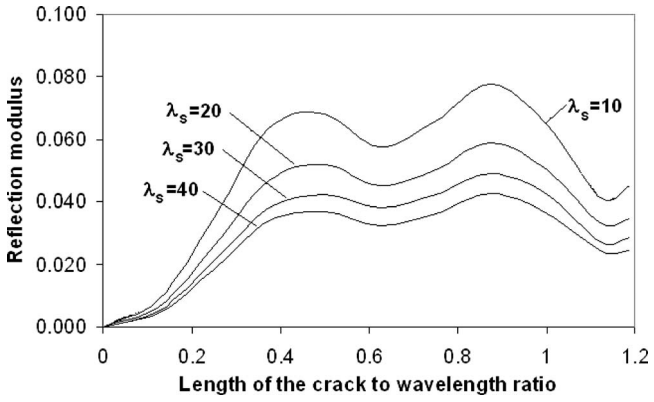


FIG. 11. FE predictions of SH_0 reflection modulus as a function of crack length for various source distances λ_S from the crack. The distance between the source and monitoring point was always $5\lambda_{SH0}$.

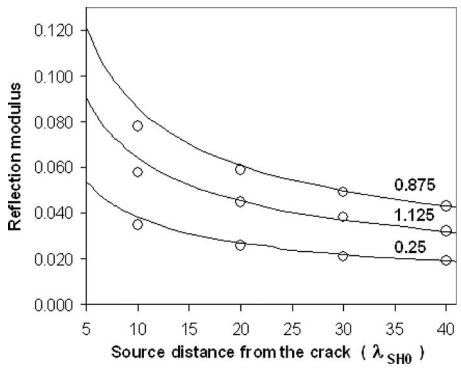


FIG. 12. FE predictions of the dependence of reflection modulus on the distance between the source and the crack for different crack length to wavelength ratios. Reflection modulus extrapolated from the source distance $40\lambda_{SH0}$ (line), reflection modulus obtained using the source distances 10, 20, and $30\lambda_{SH0}$ (empty circle).

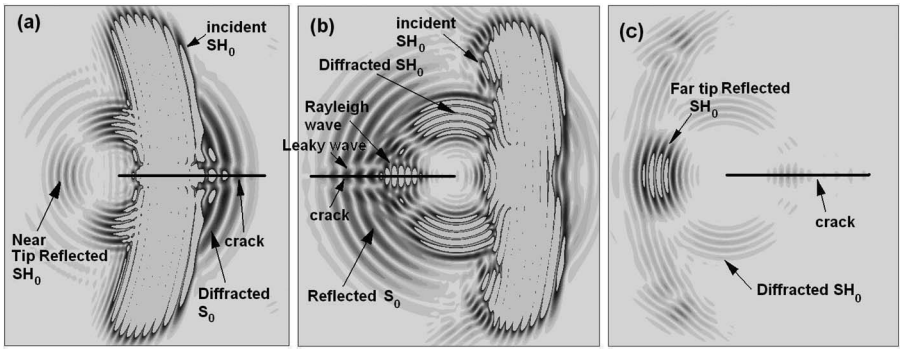


FIG. 13. Various FE time records showing the displacement amplitudes of the wave field around the crack during the interaction of the SH_0 mode with the crack. (a) Incident SH_0 mode interaction with the near tip of the crack. (b) Rayleigh type wave interaction with the far tip of the crack. (c) Reflected and diffracted waves after back propagating leaky and Rayleigh type wave interaction with the near tip of the crack.

different crack lengths to wavelength ratios 0.25, 0.875, and 1.125 is calculated using the modulus that is obtained at the source point $40\lambda_{SH0}$ from the crack, and it is extrapolated to smaller source distances and is compared with exact calculations with source distances 10, 20, and $30\lambda_{SH0}$. As we see, the results agree well when the source distance is equal or larger than $20\lambda_{SH0}$, thus confirming the compensation based on cylindrical spreading.

C. Influence of the "leaky type" surface wave on the SH_0 reflection in the incident direction

The most important signal for identifying and characterizing the crack is that which reflects directly toward the source, measured at the first point. As we saw in Sec. V A, this reflection had an undulating nature due to the presence of different surface waves. Here we take a deeper look at the generation mechanism of this reflection to explain the contribution of leaky type surface waves to the reflection. In order to understand this, we use FE simulations. We study the reflection for a very long zero-width crack $l=15\lambda_{SH0}$ and at central frequency 100 kHz. This ensures that all signals under interest can be separated and analyzed.

The wave field patterns from FE study at different time steps that describe the interaction and reflection/diffraction of the waves at the crack are shown in Figs. 13(a)–13(c). Figure 13(a) shows the wave field record when the incident SH_0 mode interacts with the front tip of the crack. It is seen that there is a small backward reflection denoted as "near tip reflected SH_0" in the figure. Due to the free boundary of the crack edges, the incident mode and also the diffraction is converted to Rayleigh type surface waves at the tip. Besides, there is also the diffracted S_0 mode which is converted to leaky type surface waves on the crack faces. All these surface waves propagate toward the end of the crack and are reflected and diffracted. This is illustrated in Fig. 13(b). We see that Rayleigh type waves are converted strongly to SH_0 modes and also to "reflected S_0" modes, which again interact with the crack face causing the leaky type surface waves to propagate. Finally, the surface waves interact with the front tip of the crack and the "far tip reflected SH_0" mode is generated, shown in Fig. 13(c). It is clearly seen that this reflection is much stronger than the near tip reflected SH_0.

The interference phenomenon between different surface waves causes the undulations in the reflected signal. As we saw, they travel with different speeds along the crack face and their phase changes are therefore also different. By predicting the wave number of these waves we can explain the occurrence of the undulations more quantitatively. Following a procedure similar to that in earlier work,[12] the wave numbers of the surface waves can be calculated by measuring the phase difference between the diffracted and reflected signals as a function of the crack length. Here we investigate the diffraction at the second point and reflection at the first point [Fig. 3(a)] at an equal distance from the crack again. Figures 14(a) and 14(b) show the corresponding time histories recorded at these points. We measure the phase ϕ_1 of the diffracted SH_0 mode (the last signal) in Fig. 14(a) by taking the Fourier transform of the gated signal. Similarly we calculate the phase ϕ_2 of the far tip reflected SH_0 mode from the Rayleigh type wave contribution and phase ϕ_3 of the re-

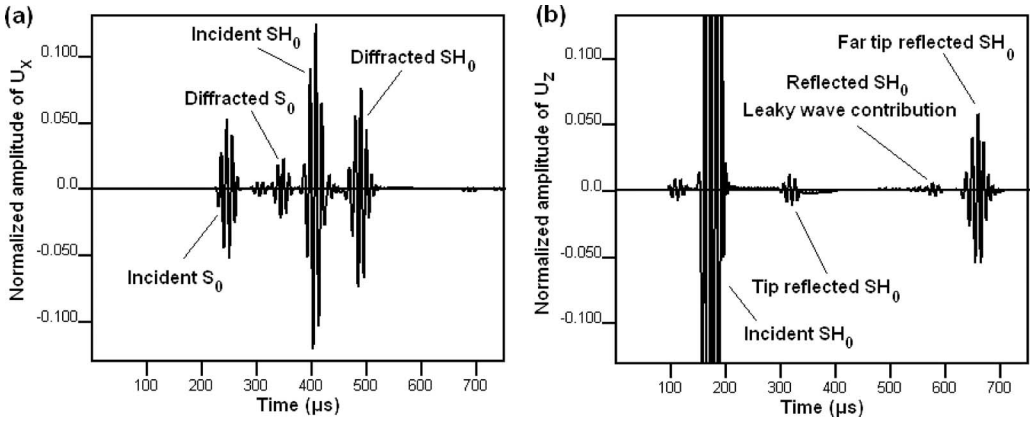


FIG. 14. FE time traces; the length of the crack is $15\lambda_{SH0}$; center frequency is 100 kHz. (a) Measurement at the second point; (b) measurement at the first point.

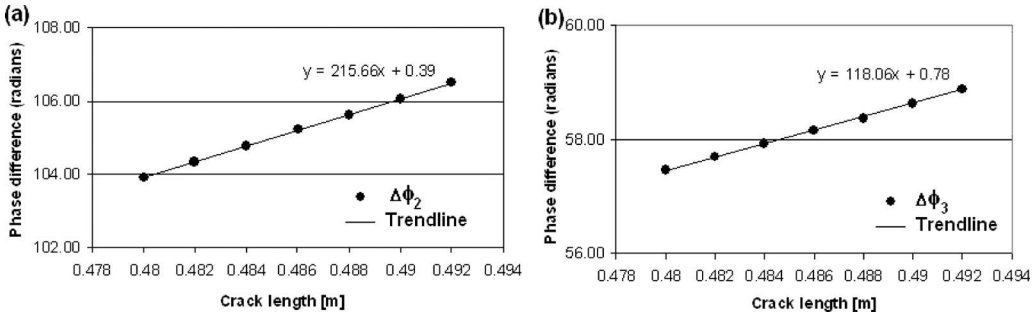


FIG. 15. Phase differences from FE simulations as a function of crack length. (a) $\Delta\phi_2$ and its slope showing the wave number of Rayleigh type surface waves; (b) $\Delta\phi_3$ and its slope showing the wave number of leaky type surface waves.

flected SH_0 from leaky type wave contribution, both shown in Fig. 14(b). Thereafter we plot the phase differences $\Delta\phi_2 = \phi_1 - \phi_2$ and $\Delta\phi_3 = \phi_1 - \phi_3$ for increasing crack lengths. The slope of these linear variations gives then the wave number for the surface wave. Figure 15(a) shows that the wave number for Rayleigh type surface waves is $k = 215.7$ rad/m. The corresponding phase velocity 2913 m/s is close to the Rayleigh wave phase velocity 2966 m/s for the material properties used in the FE simulations. Figure 15(b) gives the wave number for leaky type surface waves $k = 118.1$ rad/m, indicating that its phase velocity 5320 m/s is slightly lower than the phase velocity of the S_0 mode 5404 m/s of the studied plate.

Let us return now to Fig. 9 in Sec. V A where the reflection modulus was characterized by undulations. Here we give the explanation to the undulations, which take place for crack lengths longer than $2\lambda_{SH0}$. We are interested in the period of these undulations. We see that the reflections caused by surface waves are in the same phase when the crack length is $2.8\lambda_{SH0}$. The changing part of the phase difference $\Delta\phi_2 - \Delta\phi_3 = 97.6\,x$ gives us the phase delay between these two reflections. The first minimum occurs after a half cycle calculated as $\pi/97.6$ which corresponds to $1\lambda_{SH0}$ and the first maximum occurs after 1 cycle at $2\pi/97.6$ which corresponds to $2\lambda_{SH0}$. Similar values between the peaks and troughs are seen also in the figure; maxima occurring at 2.8, 4.8, and $6.8\lambda_{SH0}$; minima lying at 3.8 and $5.8\lambda_{SH0}$. This supports the fact that undulations in the reflected signal are caused by the interference of these different surface waves.

VI. CONCLUSIONS

FE studies, supported by experimental measurements, were performed in order to investigate the nature of the interaction of the low frequency SH_0 mode with a narrow through-thickness crack in a plate oriented in the wave propagation direction. Good agreement between the FE results and measurements was achieved. The scattered field at the crack consisted of the SH_0 modes, which were reflected in the incident direction and diffracted at 90° to that direction. The strongest reflection came from the far tip of the crack but was only seen separately in cases of long cracks. Additionally, we observed mode converted S_0 modes in the scattered field that had less energy than the SH_0 modes.

It has been shown that there were two clearly different behaviors in the reflection character depending on the length of the crack. When the length of the crack was smaller than a wavelength of the incident signal, the intervals between measured maxima and minima in the reflection modulus were smaller than a wavelength of the incident wave. This was due to the interference of the reflections coming from the near and far tip of the crack. The larger delayed far tip reflection was caused by the conversion of the incident SH_0 mode into Rayleigh type surface waves that propagated along the crack faces and radiated at the tips of the crack into the plate. For short cracks, the near tip and far tip reflections were superimposed causing an undulating behavior in the resulting reflection.

The examination of the reflection from cracks longer than a wavelength of the incident signal revealed a third reflection component caused by radiation of leaky type surface waves. These were induced by the S_0 mode interaction with the free faces of the crack and were diffracted at the far tip. It has been shown that these surface waves traveled at a higher speed than Rayleigh type surface waves and therefore also caused oscillating behavior in the reflection modulus. However, this time the interval between peaks and troughs in delayed reflection was bigger than a wavelength of the incident mode. Therefore, we observed two scattering regimes: one describing reflections at short cracks and the other at long cracks. The range of the two regimes depends on the chosen time length of the incident signal.

ACKNOWLEDGMENTS

The support from Estonian Science Foundation Grant No. 6169 and Estonian Ministry of Education and Science Grant No. SF0140072s08 is acknowledged.

[1] P. Wilcox, "Omni-directional guided wave transducer arrays for the rapid inspection of large areas of plate structures," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 699–709 (2003).
[2] D. N. Alleyne and P. Cawley, "Long range propagation of Lamb waves in chemical plant pipework," Mater. Eval. **55**, 504–508 (1997).
[3] P. Wilcox, M. Lowe, and P. Cawley, "Lamb and SH wave transducer arrays for the inspection of large areas of thick plates," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 2000), Vol. 19, pp. 1049–1056.
[4] M. J. S. Lowe and O. Diligent, "Low-frequency reflection characteristics of the s_0 Lamb wave from a rectangular notch in a plate," J. Acoust. Soc. Am. **111**, 64–74 (2002).
[5] M. J. S. Lowe, P. Cawley, J.-Y. Kao, and O. Diligent, "The low frequency reflection characteristics of the fundamental antisymmetric Lamb wave a_0 from a rectangular notch in a plate," J. Acoust. Soc. Am. **112**, 2612–2622 (2002).

[6]D. N. Alleyne, B. Pavlakovic, M. J. S. Lowe, and P. Cawley, "Rapid long-range inspection of chemical plant pipework using guided waves," Insight **43**, 93–101 (2001).
[7]C. M. Fortunko and R. B. King, "Nondestructive evaluation of planar defects in plates using low-frequency shear horizontal waves," J. Appl. Phys. **53**, 3450–3458 (1982).
[8]Z. Abduljabbar and S. K. Datta, "Diffraction of horizontally polarized shear waves by normal edge cracks in a plate," J. Appl. Phys. **54**, 461–472 (1983).
[9]M. Koshiba, K. Hasegawa, and M. Suzuki, "Finite-element solution of horizontally polarized shear wave scattering in an elastic plate," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **34**, 461–466 (1987).
[10]J. Ditri, "Some results on the scattering of guided elastic SH waves from material and geometric waveguide discontinuities," J. Acoust. Soc. Am. **100**, 3078–3087 (1996).
[11]A. Demma, P. Cawley, and M. J. S. Lowe, "Scattering of the fundamental shear horizontal mode from steps and notches in plates," J. Acoust. Soc. Am. **113**, 1880–1891 (2003).
[12]R. Rajagopal and M. J. S. Lowe, "Short range scattering of the fundamental shear horizontal guided wave mode normally incident at a through-thickness crack in an isotropic plate," J. Acoust. Soc. Am. **122**, 1527–1538 (2007).
[13]D. Alleyne and P. Cawley, "The interaction of Lamb waves with defects," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 381–397 (1992).
[14]P. Fromme and M. B. Sayir, "Measurement of the scattering of a Lamb wave by a through hole in a plate," J. Acoust. Soc. Am. **111**, 1165–1170 (2002).
[15]M. Castaings, E. Le Clezio, and B. Hosten, "Modal decomposition method for modeling the interaction of Lamb waves with cracks," J. Acoust. Soc. Am. **112**, 2567–2582 (2002).
[16]I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1970).
[17]J. A. Ogilvy and J. A. G. Temple, "Diffraction of elastic waves by cracks: Application to time-of-flight inspection," Ultrasonics **21**, 259–269 (1983).
[18]K. F. Graff, *Wave motion in Elastic Solids* (Dover, New York, 1975).
[19]B. Pavlakovic, M. Lowe, D. Allayne, and P. Cawley, "DISPERSE: A general purpose program for creating dispersion curves," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 1997), Vol. 16, pp. 185–192.
[20]D. N. Alleyne and P. Cawley, "Optimization of Lamb wave inspection techniques," NDT & E Int. **25**, 11–22 (1992).
[21]A. Demma, P. Cawley, and M. J. S. Lowe, "The reflection of the fundamental torsional mode from cracks and notches in pipes," J. Acoust. Soc. Am. **114**, 611–625 (2003).
[22]ABAQUS v6.5, Analysis User's Manual (2004).
[23]M. Castaings and C. Bacon, "Finite element modeling of torsional wave modes along pipes with absorbing materials," J. Acoust. Soc. Am. **119**, 3741–3751 (2006).
[24]P. Wilcox, "Modeling the excitation of Lamb and SH waves by point and line sources," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 2003), Vol. **23**, pp. 206–213.
[25]K.-J. Bathe, *Finite Element Procedures in Engineering Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1982).
[26]A. A. Oliner, "Waveguides for acoustic surface waves: A review," Proc. IEEE **64**, 615–627 (1976).
[27]C. T. Schroder and W. R. Scott, Jr., "On the complex conjugate roots of the Rayleigh equation: The leaky surface wave," J. Acoust. Soc. Am. **110**, 2867–2877 (2001).

Influence of material viscoelasticity on the scattering of guided waves by defects

Mihai V. Predoi[a)]
Department of Mechanics, University Politehnica Bucharest, Splaiul Independentei 313, Sector 6, Bucharest 060042, Romania

Michel Castaings and Ludovic Moreau
Laboratoire de Mécanique Physique, UMR CNRS 5469, Université Bordeaux 1, 351 cours de la Libération, 33400 Talence, France

(Received 5 February 2008; revised 29 May 2008; accepted 23 June 2008)

Guided waves are potential candidates for the nondestructive evaluation of viscoelastic structures due to their relatively long range of propagation. The major drawback is the difficulty in interpreting the scattered waves especially at high frequency-thickness values since many modes then exist. Moreover, in damping material waveguides, each mode of the scattered field has its own attenuation. Viscoelastic material characterization has been widely investigated by many authors in the past, but very few are treating multimodal scattering by discontinuities in viscoelastic guide. The scattering of a pure fundamental mode incident on a trough in a viscoelastic plate is investigated in this paper, over a relatively large frequency range, with up to five scattered modes. A hybrid two-dimensional finite element and modal projection method is used, based on modal orthogonality, to obtain the relative energy fluxes of each mode. Experiments are also made to validate the numerical predictions. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977604]

PACS number(s): 43.35.Mr, 43.20.Mv, 43.35.Cg [RLW] Pages: 2883–2894

I. INTRODUCTION

Guided waves have been extensively studied in the past decades for their potential use in long-range detection of various defects in structures. The wide spreading of composite structures has opened the research field of wave scattering in viscoelastic materials. The difficulties associated with the rheological model to be implemented in the dispersion curve routines are doubled by the difficulties in modal analysis of a decaying scattered field.

Many researches have been conducted in the field of viscoelastic material characterization[1,2] but quite few in the theoretical or numerical analysis of acoustic fields scattered by defects in such materials. The influence of high material damping on Lamb wave dispersion behavior has been investigated, and specific significant effects on the mode propagation have been shown.[3] These results motivate our interest in studying scattering phenomena in absorbing materials.

Recently, constitutive relations have been solved in the Fourier domain using the finite element (FE) method, and the dynamic responses of viscoelastic bars or plates to either thermal or mechanical loads have been predicted by considering complex elasticity coefficients as input data.[4] The same formalism is used in the present paper for investigating a scattering problem in a viscoelastic material plate. Moreover, a specific orthogonality relation-based postprocessing technique is used for the analysis of the simulated scattered fields.[5] A similar approach has been developed for optimizing health monitoring systems based on Lamb wave propagation.[6] Also based on a modal decomposition method, a theoretical study of Lamb mode scattering by a finite-length delamination in an absorptive composite plate has been proposed.[7] The effect of viscoelastic coating materials on the inspection range using guided wave has also been investigated, and adequate frequency range was determined by a hybrid FE boundary element method, which explicitly includes the attenuating characteristics of the coating. Through parametric studies some trends and features were highlighted that could be used for remote characterization of defects.[8] More recently, the influence of material viscoelasticity on Lamb wave scattering has been studied for the case of a thin notch in a Perspex (polymethylmethacrylate) plate.[9] The amount of energy converted among the fundamental A_0 and S_0 modes has been numerically and experimentally determined.

The present study extends the researches done by other authors to the multimodal scattering by a trough in a viscoelastic plate. The incident mode is scattered as reflected and transmitted fields, each one made of several propagating modes, with different levels of attenuation. The viscoelastic material makes each mode propagate with amplitude decay according to the imaginary part of the wavenumber at a specific frequency. The modes are chosen to be called *quasipropagating* if the ratio between imaginary and real parts of their wavenumbers is lower than 10%, a value which allows experimental detection of waves at relatively large distances. In this analysis, the transition from highly attenuated to quasipropagating modes is also investigated by changing the frequency of the incident wave. Another point of interest in the study is the effect of the notch width, which allows resonances, with direct influence on the reflection and transmis-

[a)]Author to whom correspondence should be addressed. Electronic mail: predoi@cat.mec.pub.ro

TABLE I. Measured viscoelastic moduli for Perspex (in gigapascals).

C^*_{ij}	$C^*_{11}=C^*_{22}=C^*_{33}$	$C^*_{12}=C^*_{13}=C^*_{23}$	$C^*_{44}=C^*_{55}=C^*_{66}$
$C'_{ij}(1+\tan\delta I)$	$7.73\times(1+0.026I)$	$4.37\times(1+0.032I)$	$1.77\times(1+0.043I)$

sion coefficients. The well-known semianalytical finite element (SAFE) method is used to establish the dispersion curves and through-thickness fields of the modes.[10–12] The FE method is used to predict the scattered ultrasonic field in the frequency domain, and the principle of orthogonality between Lamb modes is applied to determine the spectral amplitude of each scattered mode. The energy balance is calculated to verify the accuracy of the results. Experiments are finally conducted to check the numerical results over a certain range of frequencies.

II. GUIDED WAVES IN PERSPEX PLATE

The complex structures of modern composites require advanced and efficient techniques to solve for the dispersion curves of viscoelastic materials. Several methods now exist, and dispersion curves for multilayered, anisotropic, viscoelastic plates can be obtained for wide frequency ranges.[10–17] In the present paper, the SAFE method is used with complex elastic constants for modeling viscoelastic materials. These constants are expressed in a Cartesian frame with axis 1 normal to the plate and with axis 2 coinciding with the direction of propagation. The dispersion curves can be plotted as phase velocity or attenuation versus frequency from the complex wavenumbers supplied by the SAFE method:

$$k = k' + Ik'', \tag{1}$$

where k'' is directly related to modal attenuation and $I=\sqrt{-1}$.

A 3 mm thick Perspex plate of mass density 1.45 g/cm^3 has been first characterized using a standard immersion technique[2] in the 0.1–1 MHz frequency range. In Ref. 2 it has been shown that the attenuations of either longitudinal or shear bulk waves in polymer-based materials vary linearly with frequency in a given frequency range roughly extending from 0.1 to 3 MHz. Consequently, in this range, the viscoelastic moduli can be assumed to be frequency independent. The resulting viscoelastic coefficients are given in Table I. These moduli are complex quantities, the real parts representing the material stiffness and the imaginary parts its viscoelasticity; $\tan\delta$ is the ratio between these imaginary and real parts. All other C^*_{ij} besides those given in Table I are null due to the symmetry of the material.

The dispersion curves are presented in terms of phase velocity and attenuation for the first five modes of the plate in two cases: (1) neglecting the viscoelasticity by canceling the imaginary parts of the viscoelastic coefficients [Figs. 1(a) and 1(c)] and (2) taking into account the material viscoelas-

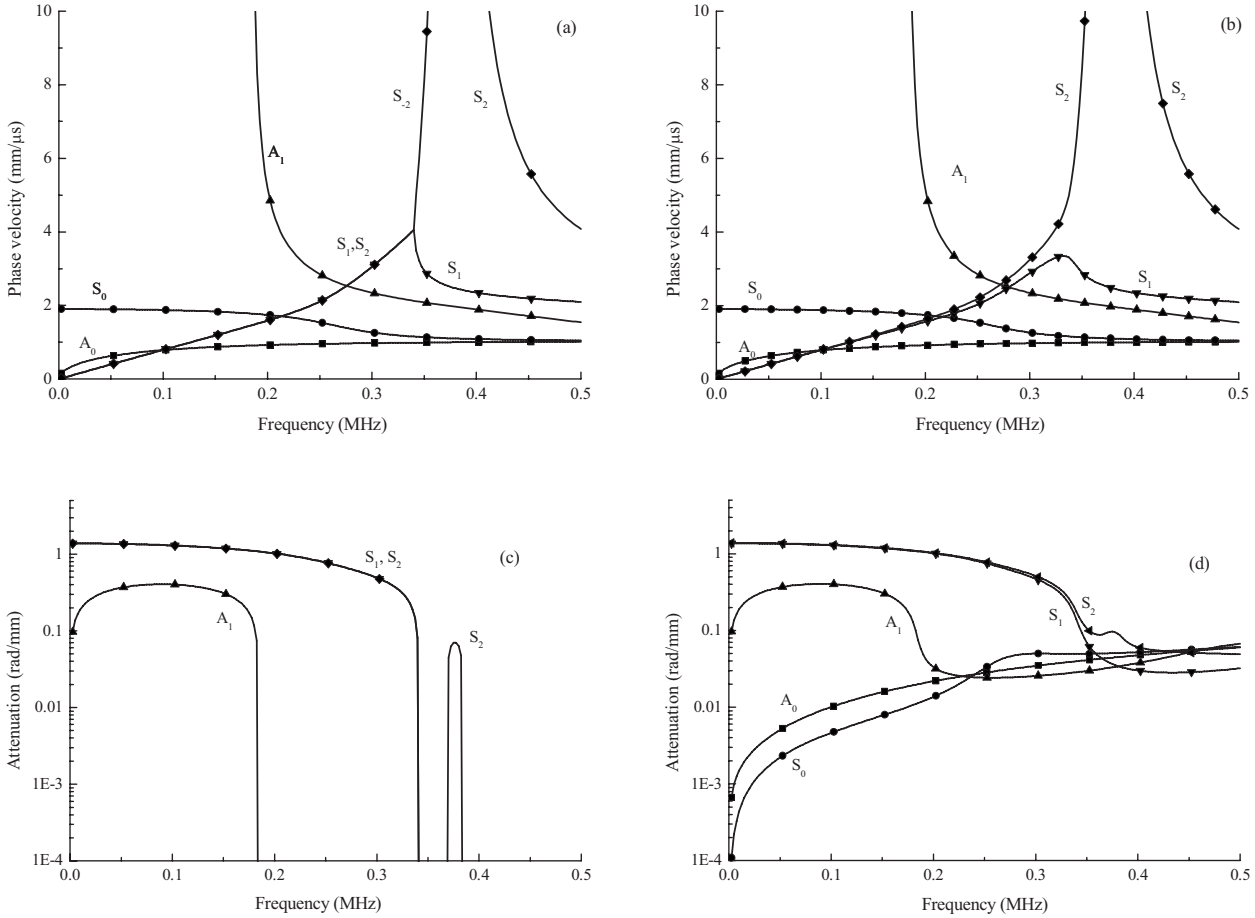


FIG. 1. Comparison between phase velocities (top) and attenuations (bottom) for the elastic (left) and viscoelastic (right) cases.

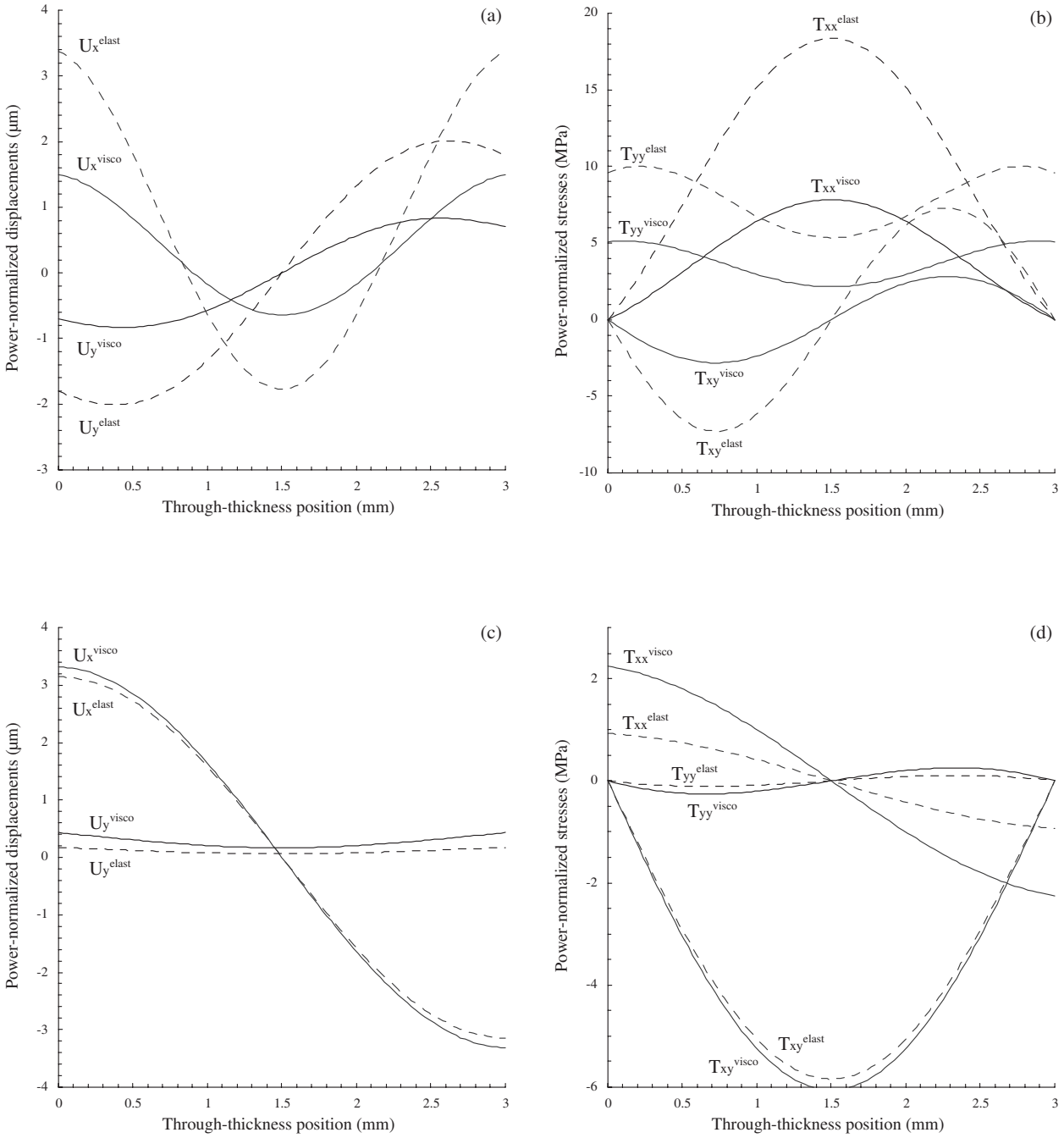


FIG. 2. Comparison between modal displacements (left) and modal stresses (right) for S_1 at 0.34 MHz (top) and A_1 at 0.18 MHz (bottom); predictions with viscoelastic (—) and elastic (- -) properties.

ticity [Figs. 1(b) and 1(d)]. The phase velocities are almost unaffected by the imaginary parts of the coefficients. An exception is a narrow frequency interval below the cutoff of the S_1 mode (approximately 0.35 MHz), for which the viscoelastic material presents two distinct curves, whereas in the elastic case the S_1 cutoff point is a common value for the S_1 and S_2 modes, after which the S_2 mode becomes an inverse mode (phase and group velocities with opposite signs), noted S_{-2} in Fig. 1(a), followed by an imaginary branch continued with real values (back to S_2 mode), which are almost coincident with the real parts of the velocities for the viscoelastic plate. In this paper, modes with attenuations lower than 0.5 dB/mm (0.057 rad/mm in Fig. 1) are considered as quasipropagating, meaning that they can be detected until a sufficiently long distance from the generation zone with appropriate equipment.

Figure 2 presents the in-plane (U_x) and out-of-plane (U_y) displacement components, as well as stress components (T_{xx}, T_{yy}, and T_{xy}) for the S_1 and A_1 modes at 0.34 and 0.18 MHz, respectively. These modes and frequencies have been selected to emphasize the effect of viscoelasticity on the through-thickness mode shapes. Not shown here, although quite interesting is the phase differences between these various components, which are close but not equal to 0 or 90 deg for the viscoelastic material, as it is the case for elastic ma-

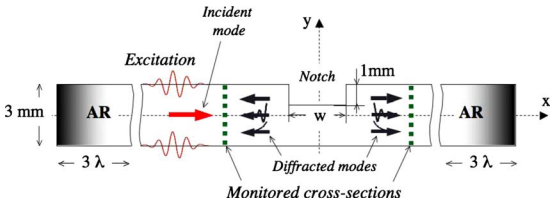


FIG. 3. (Color online) Schematic of the FE model including the trough and other specific regions.

terial guides. It can therefore be expected that these unconventional mode shapes produce significant differences between scattering phenomena produced in viscoelastic and elastic materials as the frequency increases, and this is specifically the focus of this paper.

III. SIMULATIONS AND POSTPROCESSING

A. FE model

The aim of this study is to demonstrate the effect of material viscoelasticity on the scattered mode amplitude and energy flux. In this purpose, the scattering of a pure Lamb mode incident on a trough in a Perspex plate is considered. In the numerical simulations, two situations will be alternatively investigated: one considering the plate as a viscoelastic medium and the other considering it as an elastic medium. This will allow some quantification of the effects of the material viscoelasticity. Since the material is very slightly anisotropic, Lamb waves are supposed to produce significant strains essentially in their plane of propagation, which is formed by the Ox axis, i.e., the direction of propagation, and the Oy axis, which is normal to the plate (Fig. 3). Therefore, a two-dimensional (2D) model with plane strain conditions is appropriate for this study.

For an orthotropic linearly viscoelastic material of density ρ, the displacement 2D general equation is written as follows in the frequency domain:

$$Q_{11}^{*}\frac{\partial^2 \tilde{u}_x}{\partial x^2}+Q_{66}^{*}\frac{\partial^2 \tilde{u}_x}{\partial y^2}+(Q_{12}^{*}+Q_{66}^{*})\frac{\partial^2 \tilde{u}_y}{\partial x\,\partial y}+\rho\omega^2\tilde{u}_x=0,$$

$$Q_{22}^{*}\frac{\partial^2 \tilde{u}_y}{\partial y^2}+Q_{66}^{*}\frac{\partial^2 \tilde{u}_y}{\partial x^2}+(Q_{21}^{*}+Q_{66}^{*})\frac{\partial^2 \tilde{u}_x}{\partial x\,\partial y}+\rho\omega^2\tilde{u}_y=0, \quad (2)$$

where $(\tilde{u}_x,\tilde{u}_y)$ is the Fourier transform of the displacement vector, ω is the circular frequency, and indices 1 and 2 are used for defining the parameters Q_{ij}^{*} or C_{ij}^{*} with 1 and 2 corresponding to the x and y directions, respectively. With plane strain state ($\varepsilon_{xz}=\varepsilon_{yz}=\varepsilon_{zz}=0$), the coefficients Q_{ij}^{*} are related to the components C_{ij}^{*} of the complex stiffness matrix by the very simple relation $Q_{ij}^{*}=C_{ij}^{*}=C_{ij}'=IC_{ij}''$, as explained in Ref. 4. To be implemented in the FE code[18] used in this study, the above differential equations must be written in the following form:

$$\nabla\cdot(c\,\nabla u-\gamma)-au=0. \quad (3)$$

Some derivations allow showing that the coefficient c is a 2×2 matrix where the components are also 2×2 matrices:

$$c_{11}=\begin{pmatrix}Q_{11}^{*} & 0\\ 0 & Q_{66}^{*}\end{pmatrix},\quad c_{12}=\begin{pmatrix}0 & Q_{12}^{*}\\ Q_{66}^{*} & 0\end{pmatrix},$$

$$c_{21}=\begin{pmatrix}0 & Q_{66}^{*}\\ Q_{21}^{*} & 0\end{pmatrix},\quad c_{22}=\begin{pmatrix}Q_{66}^{*} & 0\\ 0 & Q_{22}^{*}\end{pmatrix}. \quad (4)$$

The matrix a can be identified as

$$a=\begin{pmatrix}-\rho\omega^2 & 0\\ 0 & -\rho\omega^2\end{pmatrix} \quad (5)$$

and γ is a null matrix.

The boundary conditions are stress-free, except for two small domains facing each other, on the upper and lower surfaces of the plate. These are 10 mm wide and centered at $x=-15$ mm and are submitted to normal stresses defined as sinusoidal functions of the product $k'\times x$, k' being the wavenumber of the desired incident mode, modulated in amplitude by a Gaussian window. They constitute the excitation required to launch a pure (single) incident Lamb wave, and the absence of other incident modes is tested using the orthogonality formula (6). Unwanted reflections from the left- and right-hand sides of the domain are eliminated by using absorbing regions, which consist in the progressive increase in the imaginary parts of the viscoelastic coefficients.[4,5,9] Between both absorbing regions, the model includes the excitation and the trough, as well as small domains in which the resulting stresses and displacements are monitored across the thickness of the plate, at two positions along the x-axis. The first position is located 1 mm before the discontinuity (to the left where incident and reflected modes coexist) and the second one is 1 mm past the discontinuity (to the right where transmitted modes are produced). The simulation is run twice over 61 frequency values comprised between 0.2 and 0.35 MHz, once by simulating the material as an elastic medium, and then as a viscoelastic one. Due to the asymmetry of the trough with respect to the middle plane of the plate, modal conversion between modes of different symmetries will occur.

The overall length of the FE model is 400 mm. Each absorbing region is extended over 140 mm long, which corresponds to three times the maximum wavelength of the modes in the frequency range of investigation (i.e., that of mode A_1 at 0.2 MHz). A number of 35 962 quadratic elements with a total of 147 774 degrees of freedom have been used. Separate analyses for S_0 or A_0 incident modes are effectuated in Secs. IV B and IV C, respectively. Before studying these two cases, a series of tests has been performed in order to select a width of the trough for which the material viscoelasticity provides well visible effects on the scattered waves. This parametric study is presented in Sec. IV A, the depth of the trough and the frequency being fixed quantities.

B. Orthogonality-based postprocessing

A modal decomposition technique is used to identify in the scattered fields each mode that may propagate to the left (incident and reflected modes) and to the right (transmitted modes) sides of the trough. It is based on the well-known

reciprocity formula[19] between any different Lamb modes m and n. A similar technique described in detail in Ref. 5 uses the following orthogonality relation:[20]

$$\int_0^H (\sigma_{xy}^n u_y^m + \sigma_{xy}^m u_y^n - \sigma_{xx}^n u_x^m - \sigma_{xx}^m u_x^n) dy = A(n)\delta_{n,m}, \quad (6)$$

where $\delta_{n,m}$ is the Krönecker delta symbol and $A(n)$ is a normalization factor, which depends on the modal amplitude.

At each frequency and for each mode n from the dispersion curves, the stress and displacement functions of y are integrated, and the value of $A(n)$ is thus calculated. These normalization factors depend on the scale used for stresses and displacements during the solution of the dispersion equation. Next, the same orthogonality relation is applied between the scattered fields (reflected and transmitted) and each mode n, the amplitude of which is to be determined. For this purpose, the scattered fields are infinite series of Lamb modes:

$$\begin{Bmatrix} u_x \\ u_y \end{Bmatrix} = \sum_{n=1}^{\infty} S(n) \begin{Bmatrix} u_x^n \\ u_y^n \end{Bmatrix}, \quad \begin{Bmatrix} \sigma_{xx} \\ \sigma_{xy} \end{Bmatrix} = \sum_{n=1}^{\infty} S(n) \begin{Bmatrix} \sigma_{xx}^n \\ \sigma_{xy}^n \end{Bmatrix}, \quad (7)$$

in which $S(n)$ are the modal amplitudes. Practically in Eq. (6), m is replaced by the scattered fields and n is replaced by the mode of interest. The result of the orthogonality between the mode n and the scattered fields is then noted $B(n)$. As explained in Ref. 5, the amplitude of the mode n is finally given by the quotient between $B(n)$ and $A(n)$: $S(n) = B(n)/A(n)$. A reasonable number of modes (3,…,5 for the frequency range used in the following) in decreasing order of the wavenumbers are selected from the point of view of fulfilling the energy balance. The calculation of the amplitude of the scattered modes is then effectuated at each frequency. The modal energy flux can be obtained at a given position along the waveguide and frequency $f(\omega = 2\pi f)$ as

$$W_n(\omega) = \mathrm{Im}\left[\frac{\omega S^2(n)}{2} \int_0^H [\bar{\sigma}_{xx}^n u_x^n + \bar{\sigma}_{xy}^n u_y^n] dz \right], \quad (8)$$

where $\bar{\sigma}_{xk}^n (k = x, y)$ are the power-normalized stress field components of mode n. The orthogonality formula used in this paper has the main advantage of extracting correct $B(n)$ coefficients for modes propagating in each of the two directions through the selected plate sections. It allows separation between the forward propagating modes (toward $x > 0$) and backward propagating modes (toward $x < 0$) to be made, including the nonpropagating modes scattered from the notch that have no influence on the amplitudes of the quasi-p. However, due to attenuation, the modal amplitudes once determined at one position x_1 must be corrected for another section according to

$$S_n(n)|_{x=x_2} = S_n(n)|_{x=x_1} \exp[-k_n''(x_2 - x_1)]. \quad (9)$$

The results are presented as ratios of modal energy fluxes to the incident energy flux:

$$\frac{W_{\text{scattered mode}}(\omega)}{W_{\text{incident}}(\omega)}. \quad (10)$$

From a practical point of view, the positions selected for the inspection sections are important because highly damped modes could not be extracted from the scattered fields too far from the discontinuity.

IV. LAMB WAVES SCATTERED BY A TROUGH

A. Selection of a trough width

In this study, the frequency was chosen to be equal to 0.34 MHz because the influence of material viscoelasticity on displacement and stress fields is very pronounced at this frequency, as mentioned in Sec. II and as shown in Fig. 2. The trough is 1 mm deep, and its width, noted w, is defined as a variable parameter in this preliminary model, with values increasing from 0 to 6 mm, every 0.5 mm. In order to simulate a varying width, the material properties are defined as space functions. At the location of the trough, the density and the viscoelastic moduli of the material are divided by a factor equal to 10^{30} so that their values tend toward zero. The absorbing regions described in Sec. III A are used in this model. However, a finer mesh than that mentioned in Sec. III A has been used in the whole area supposed to contain the trough with varying width in order to ensure a good level of accuracy in the calculated stress fields. As a consequence, the number of degrees of freedom was equal to 240 000 for this specific model.

The orthogonality relation-based processing described in Sec. III A has been used to quantify the energy of the incident mode and that of each transmitted mode. This formula separates individual mode amplitudes so that the scattered nonpropagating modes do not influence the amplitudes of the other modes. The through-thickness displacement and stress fields have been monitored at two locations: (1) at the left-hand side of the notch to see the incident mode and (2) at the right-hand side of the notch to see the transmitted modes. These monitoring lines were located 0.01 mm distance from the notch to avoid amplitude corrections due to the viscoelasticity out of the notch. In the model, the distance between the source and the left-hand side of the notch was kept constant to have always the same incident energy for all values of w. The results of this preliminary study are presented in Fig. 4, which shows the transmitted energy ratios, calculated at the right edge of the notch when the A_0 mode is incident. The study reveals that, except for the S_0 mode, the differences between the ratios calculated in the elastic or viscoelastic materials are well pronounced when the width of the trough becomes greater than 2 mm with a peak for the A_0 mode at $w = 4$ mm. Therefore, in both following applications, the width of the trough is chosen to be equal to 4 mm. However, it is interesting to note that the differences between the two sets of simulations (elastic and viscoelastic cases) do not linearly increase with the width of the trough due to some resonance phenomena, which reduce these differences for some modes and some values of w, see, for instance, the transmission of A_0 for $w = 6$ mm in Fig. 4(a). These well-known effects will depend on the ratio between w and the mode wavelength.

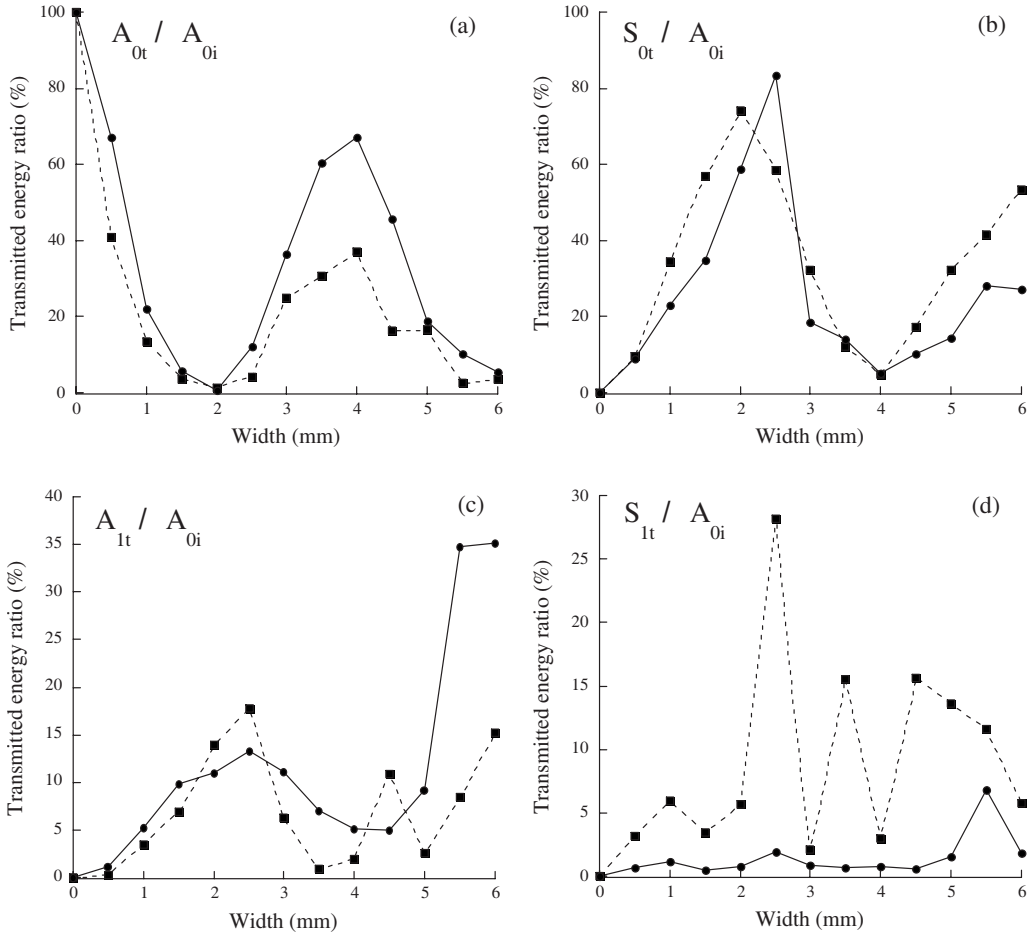


FIG. 4. Transmitted energy ratios vs trough width for modes (a) A_0, (b) S_0, (c) A_1, and (d) S_1, at 0.34 MHz, when the A_0 Lamb mode is incident on a 1 mm deep trough: (●) elastic material and (■) viscoelastic material.

B. Incident S_0 mode: Elastic and viscoelastic models for the material

The notch is now 1 mm deep and 4 mm wide, and the frequency becomes the variable parameter varying from 0.2 to 0.35 MHz. This first application concerns the incident S_0 mode. The excitation is produced by normal stresses, which are symmetrical with respect to the median plane of the plate. The scattered fields are monitored from the FE solutions, on each side of the trough and through the thickness of the plate, at distances of 5 mm remote from the center of the trough. Using the modal orthogonality, the amplitudes of the incident, reflected, and transmitted modes are obtained at these positions. Each modal amplitude is then corrected using a compensating factor that takes into account the mode attenuation over a 3 mm long propagation distance so that modal energy fluxes can be determined at each border of the trough ($x=\pm 2$ mm). Thus, only the influence of the material viscoelasticity underneath the trough is taken into account. Energy flux ratios defined as the energy of scattered waves divided by that of the incident wave at the two borders of the trough are plotted versus the frequency in Fig. 5 for the reflected modes and in Fig. 6 for the transmitted modes.

It can be remarked that the values of the reflection and transmission coefficients are affected by the viscoelasticity of the material, with a reduction in peak values superior to 50% at some frequencies [see, for example, transmitted mode A_0 at 0.29 MHz in Fig. 6(a) or reflected mode S_0 at 0.32 MHz in Fig. 5(b)]. However, it is not the case for all modes or for all frequencies. This is likely to be a consequence of the defect geometry since it is well known that the amplitudes of Lamb modes in the scattered field depend on the location and the geometry of the defect. For the considered defect, it seems that the influence of the viscoelasticity on these coefficients is stronger beyond 0.25 MHz for most modes [Figs. 5(a) and 5(c)–5(e)and Figs. 6(b)–6(d)]. This frequency in the dispersion curves corresponds to the beginning of clear separation between the elastic and the viscoelastic cases.

Note that S_1 and S_2 modes are evanescent waves below the cutoff frequency of S_1 for the purely elastic model (see Fig. 1), and consequently they do not carry energy in this frequency range.[19] Right above this cutoff frequency, the transmitted S_2 mode presents a lobe in the energy flux for the elastic case, which is clearly smoothed under the effect of the viscoelasticity [Fig. 6(e)]. It is interesting to remind that all modes propagate while dissipating energy in a viscoelastic waveguide.[21] Consequently, the contribution to scattering of nonpropagative modes in elastic guides will be different than that of the corresponding modes in viscoelastic guides. This is clearly shown in Figs. 6(d) and 6(e) for the S_1 and S_2 modes, which have significant contributions to the transmitted field in the case of the viscoelastic plate, and below the

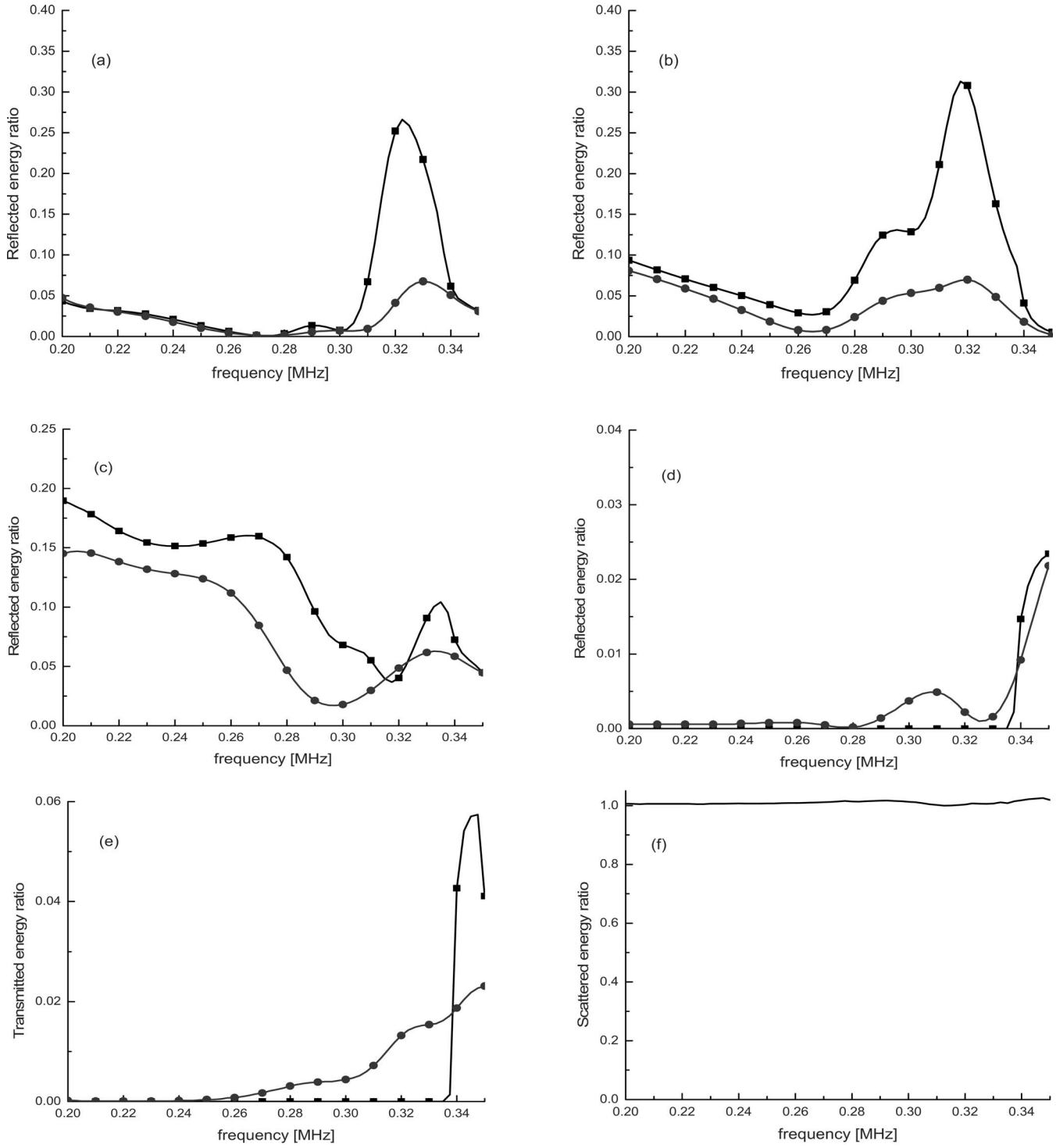


FIG. 5. Reflected energy ratios of modes (a) A_0, (b) S_0, (c) A_1, (d) S_1, and (e) S_2, relative to S_0 incident, for elastic (■) and viscoelastic (●) materials. Energy balance for the elastic material (*f*).

cutoff frequency of S_1. The material viscoelasticity induces for these modes a progressively increasing transmitted energy as the frequency increases. This phenomenon can also be observed for the reflected energy, but it is less pronounced [Figs. 5(d) and 5(e)].

The material viscoelasticity produces an expected reduction in the energy fluxes for the propagating modes. The energy balance is calculated by dividing the sum of reflected and transmitted energies by the incident energy. The complement to unity of this result represents the loss within the notch region since the various mode energy fluxes are calculated right at the beginning and end of the notch. In the elastic case [Fig. 5(f)], the energy balance is equal to unity (within a $\pm 3\%$ interval of errors, which is a commonly accepted precision for the FE method and proves a correct choice of the various numerical parameters: mesh size, absorbing regions, solver, etc). This balance equal to 1 is due to the well-known fact that no propagating mode is attenuated

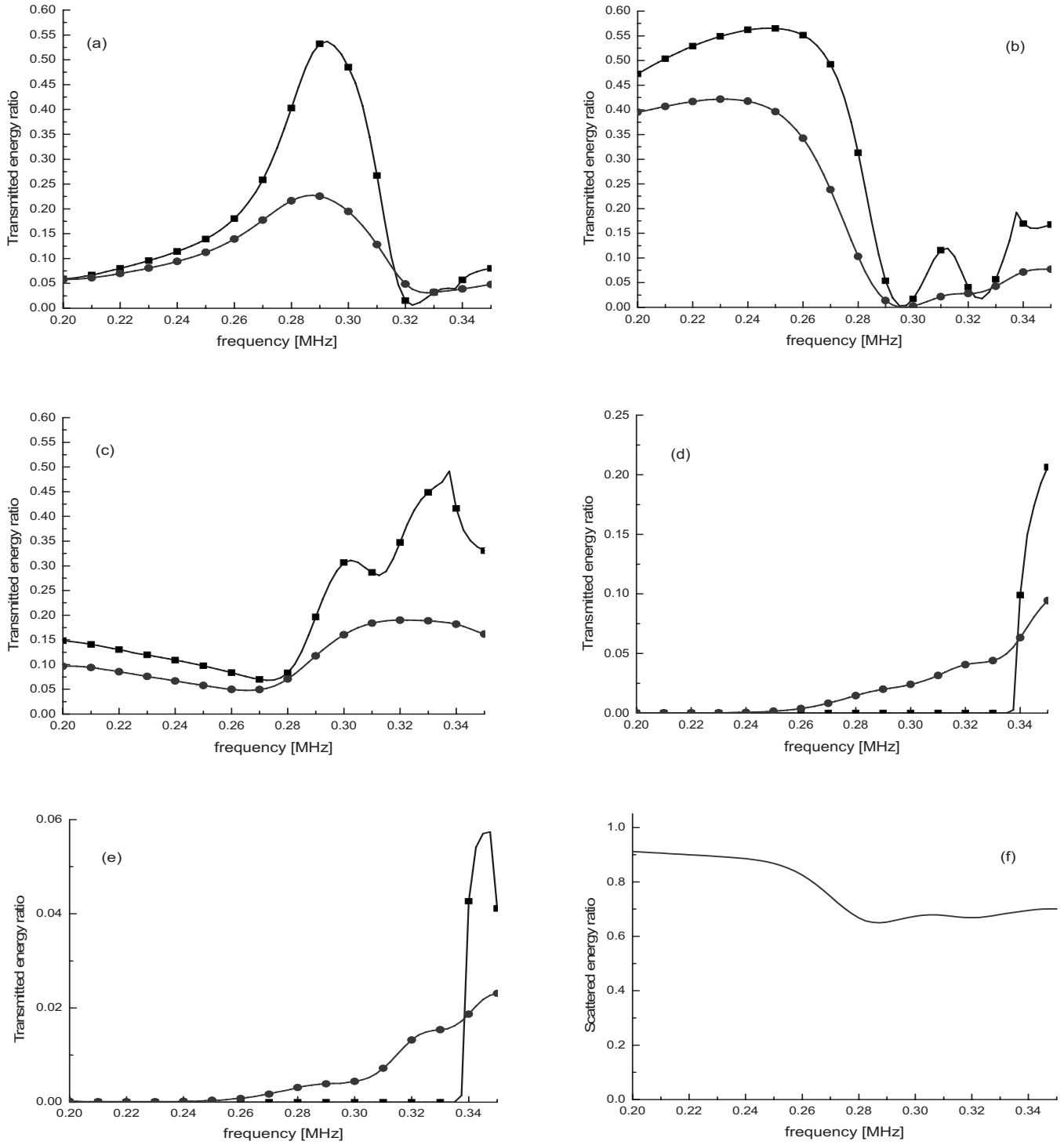


FIG. 6. Transmitted energy ratios of modes (a) A_0, (b) S_0, (c) A_1, (d) S_1, and (e) S_2, relative to S_0 incident, for elastic (■) and viscoelastic (●) materials. Energy balance for the viscoelastic material (*f*).

and no evanescent mode carries (and so dissipating) energy in elastic waveguides.[19] In the viscoelastic case, the energy balance is much smaller than unity, thus indicating non-negligible losses of energy inside the trough [Fig. 6(f)]. Two levels of energy dissipation are visible: one corresponding to about 10% loss for frequencies lower than 0.25 MHz and the other to about 30% loss at frequencies higher than 0.28 MHz. This observation is in accordance with Figs. 5(d), 5(e), 6(d), and 6(e), which show that modes S_1 and S_2 significantly contribute to the sharing of the incident mode energy, above 0.28 MHz.

C. Incident A_0 mode: Elastic and viscoelastic models for the material

The same notch, mesh, and frequency range than in the previous case were used here. The excitation, however, is set

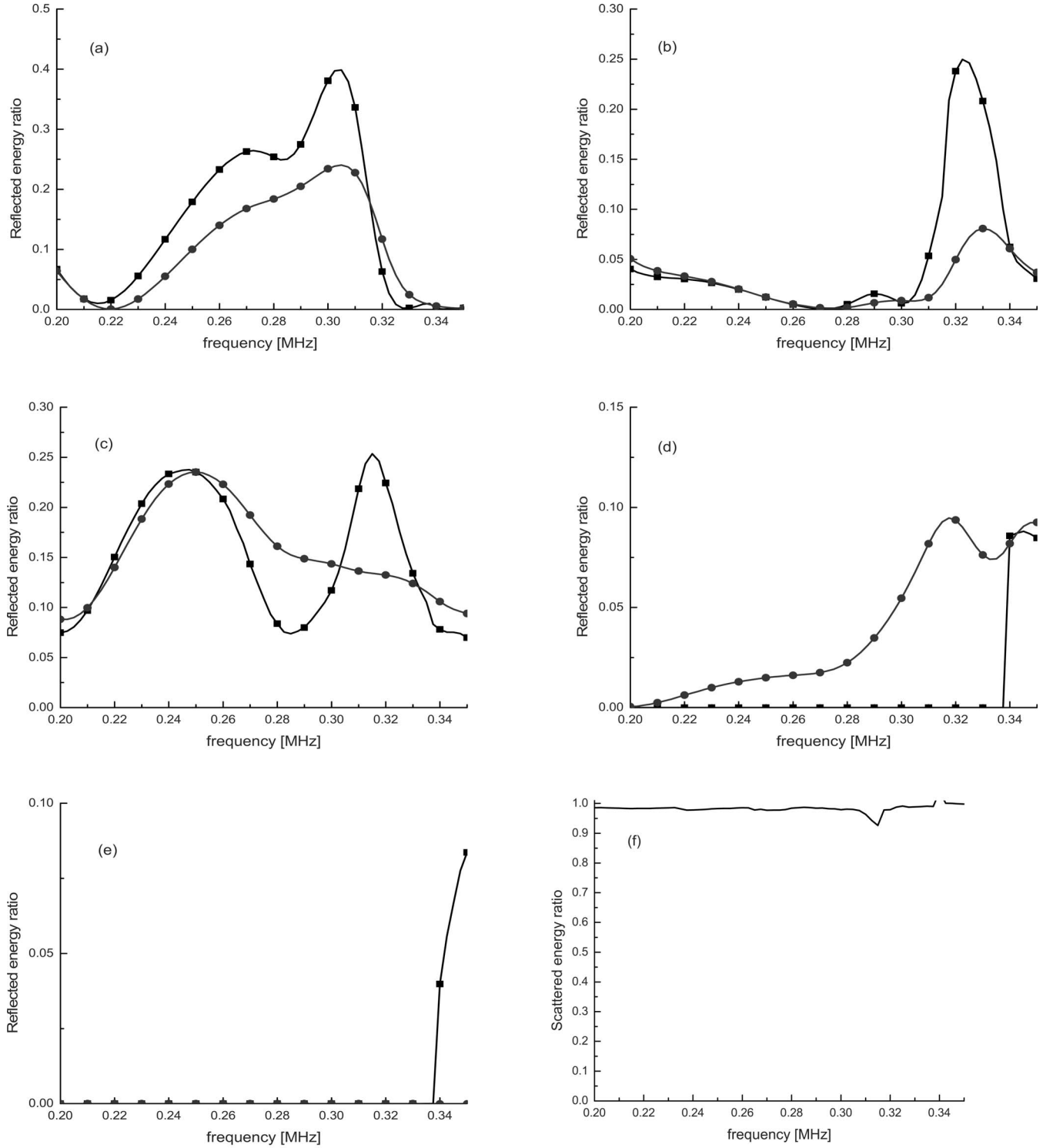


FIG. 7. Reflected energy ratios of modes (a) A_0, (b) S_0, (c) A_1, (d) S_1, and (e) S_2, relative to A_0 incident, for elastic (■) and viscoelastic (●) materials. Energy balance for the elastic material (*f*).

to launch the pure A_0 mode by applying normal stresses skew symmetrically with respect to the middle plane of the plate. Peaks of scattered energy are visible, for instance, for the reflected A_0 mode at 0.305 MHz [Fig. 7(a)] and for the A_1 mode at 0.245 MHz [Fig. 7(c)]. The viscoelastic material is shown to produce a significant attenuation of the reflected A_0 peak ($\approx -50\%$), while it has practically no effect on the reflected A_1 peak, even if the attenuation is about the same order of magnitude for these two modes at these respective frequencies [see Fig. 1(d)]. The effect on the reflected A_0 peak is likely to be coming from the very strong sensitivity to the material viscoelasticity of the displacement and stress fields of the S_1 and S_2 modes, around 0.305 MHz, as illustrated for mode S_1 at 0.34 MHz, in Figs. 2(a) and 2(b). Such changes in the through-thickness fields of some modes reveal modifications of the dynamic behavior of the medium due to the viscoelasticity, which causes in turn changes in the amplitudes of some scattered modes, even of modes having

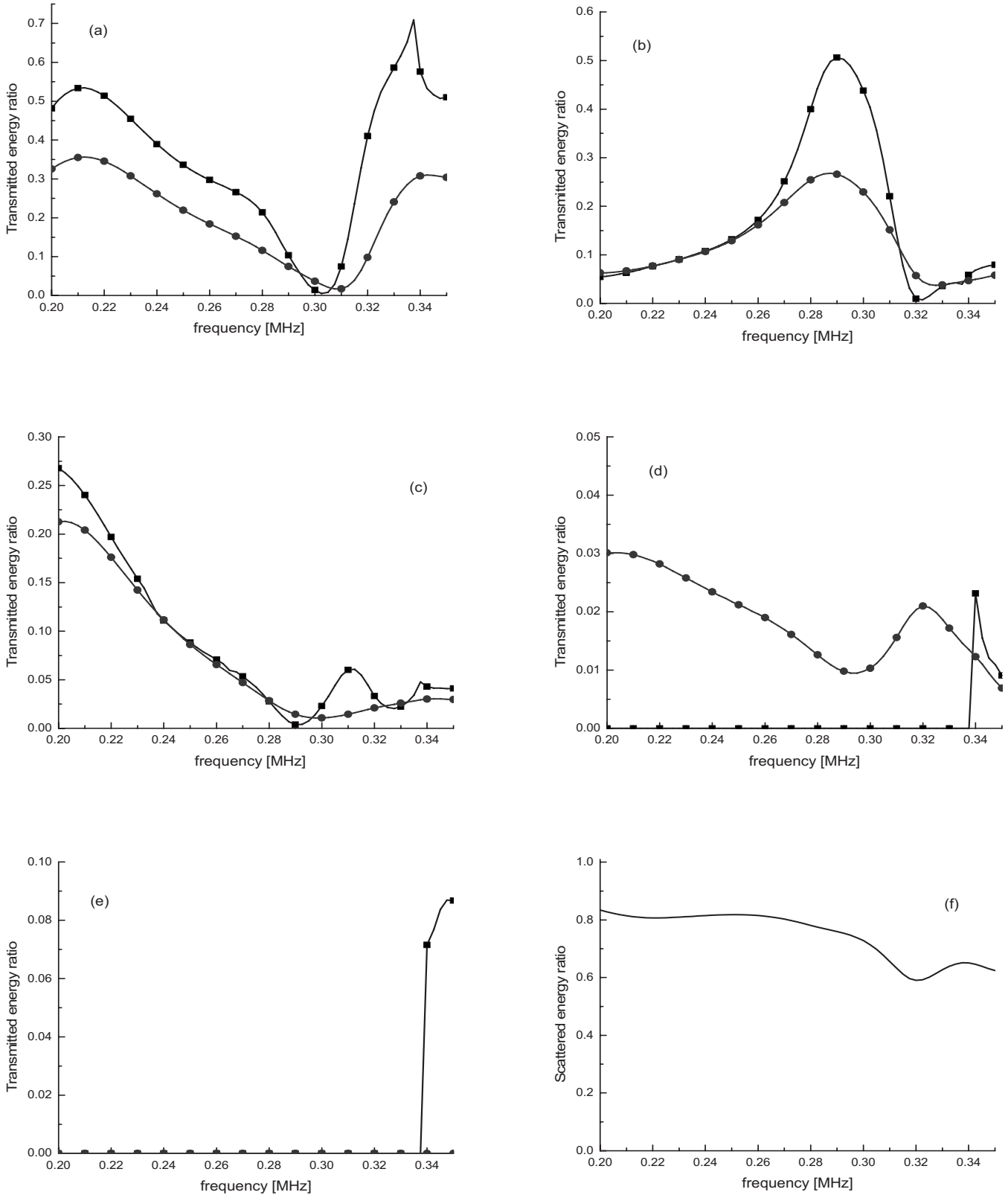


FIG. 8. Transmitted energy ratios of modes (a) A_0, (b)S_0, (c) A_1, (d) S_1, and (e) S_2, relative to A_0 incident, for elastic (■) and viscoelastic (●) materials. Energy balance for the viscoelastic material (f).

through-thickness fields not directly sensitive to the material viscoelasticity. This is precisely what is happening for the reflected A_0 mode at 0.305 MHz.

In the transmitted field, the A_0 mode reaches a first peak, at 0.21 MHz in Fig. 8(a), equal to about 54% or 35% of the incident energy in the case of the elastic or viscoelastic material, respectively. This reduction effect of the viscoelasticity on the transmission coefficient of A_0 is also confirmed for another peak close to 0.33 MHz and globally observed for all transmitted modes over the whole frequency range of inves-

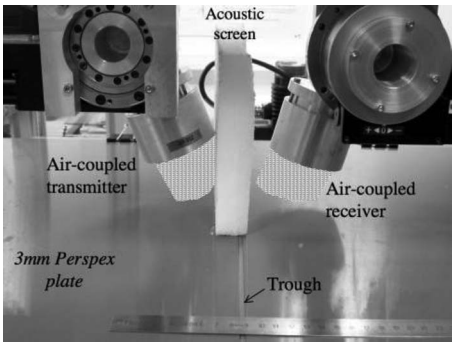


FIG. 9. Photograph of the experimental setup.

tigation. The transmitted A_1 mode, however, is slightly influenced by the viscoelasticity [Fig. 8(c)], although the reflected A_1 mode is quite modified [Fig. 7(c)]. The scattered S_0 mode has the same trend in the elastic and viscoelastic cases with peaks equal to 25% of the incident energy around 0.32 MHz in reflection [Fig. 7(b)] and to 51% around 0.29 MHz in transmission [Fig. 8(b)]. Small shifts of about 0.05 MHz toward high frequencies, for the reflected A_0, A_1, and S_0 modes, are also visible for the peaks of the energy fluxes due to the viscoelasticity of the material [Figs. 7(a)–7(c)].

Under the effect of viscoelasticity, the S_1 mode has a reflected energy flux increasing as the frequency increases, reaching a peak of 10% close to 0.31 MHz [Fig. 7(d)]. This is a significant difference with the elastic case for which S_1 does not carry any energy below 0.34 MHz (the upper limit in Fig. 7) since this is roughly its frequency cutoff. The presence of S_2 could not be detected in the scattered fields, except for the elastic case, and above its frequency cutoff, which is the same as that of S_1 [Figs. 7(e) and 8(e)].

As in the case of the incident S_0 mode, the material viscoelasticity produces an expected reduction in the energy fluxes at frequencies below approximately 0.25 MHz, for the propagating modes, more complicated influences above this frequency, and also small shifts in the frequencies of the local maxima/minima as mentioned above.

In the elastic case [Fig. 7(f)], the energy balance is again equal to unity within the numerical precision range. In the viscoelastic model [Fig. 8(f)], it is even further from unity than in the case of the incident S_0 mode. Its complement to unity indicates losses bigger than 20% of the incident energy flux for almost the whole frequency range of interest, with a maximum loss of 40%, reached at 0.32 MHz.

D. Experimental validation for the incident A_0 mode

Experiments have been made on a 3 mm thick Perspex plate with a machined 1 mm deep and 4 mm wide notch. The plate is 700 mm long and 400 mm wide to avoid any unwanted reflections from its edges, and the notch runs all along the plate width at the center of its length. This large extent of the notch allows satisfying the 2D conditions of the FE simulations. The experimental setup (Fig. 9) consists of two air-coupled circular transducers with 45 mm diameter and 0.1–0.35 MHz frequency bandwidth down to −15 dB. The frequency limits are imposed by the transducers signal to noise ratios which are acceptable only in this range. These large elements ensure a good directivity of the ultrasonic beams so that any reduction in amplitude coming from the angular aperture effect is negligible in comparison to that due to the material viscoelasticity. This was preliminary checked by comparing decreases in amplitude for different modes propagating either along an elastic glass plate or along the viscoelastic Perspex plate. One transducer is used as a transmitter and inclined at an angle of 21 deg (with respect to the normal of the plate) to launch the A_0 mode which, considering the aperture of the transducer, is the dominantly generated mode at this incidence. It is excited by a chirp temporal waveform set to have a frequency spectrum of constant amplitude (±2 dB) between 0.185 MHz and 0.38 Hz. The other transducer is successively set to appropriate angles (with respect to the normal of the plate) in order to detect either the A_0 mode (−21 deg), the A_1 mode (−7 deg), or the S_0 mode (−14 deg). These angles do not significantly vary within the frequency range of investigation. Indeed, the highest change in this angle is of 6 deg for both S_0 and A_1 modes. Considering the angular aperture of the air-coupled elements used in these experiments (a few degrees), the inclination angles of the transducers were then set to any of the coincidence angles corresponding to the center frequency.

The receiver is moved along a 50 mm in length propagation path and signals are picked up every 0.5 mm with a 100 time averaging. Four sets of 101 waveforms are thus measured: one for the incident A_0 mode, both transducers being well remote from the notch, and one for each transmitted mode, i.e., A_0, A_1, or S_0, the transmitter being sending the incident A_0 in the direction of the notch and the receiver being past the notch. The distance between the transmitter and the first position of the receiver is the same for each set of measurements, and this was chosen as small as possible to reduce the attenuation along the plate out of the notch (between 30 and 50 mm depending on the angles). An acoustic barrier was needed to avoid specular reflections from the surface of the sample. The postprocessing is based on a double (2D) fast Fourier transform, which was performed for each set of data. This supplies the amplitude versus the frequency of the normal displacement produced by the angle-selected mode at the surface of the plate and at the midposition along the displacement of the receiver, as explained in Ref. 9. These measured frequency-dependent normal displacements have then been used to quantify those at the sides of the notch (left-hand side for the incident mode and right-hand side for the transmitted modes) by considering the attenuation of each mode [Fig. 1(d)] to compensate its change in amplitude due to the propagation between the notch and the measurements position. Then, each surface normal displacement thus calculated at a notch side was divided by the corresponding (same mode and same frequency) power-normalized surface normal displacement in order to obtain the intrinsic amplitude of the mode.[20] Since the square of the intrinsic amplitude corresponds to the energy of the mode,[21] the amplitude of each transmitted mode was divided by that of the incident mode, and this ratio was squared, thus leading

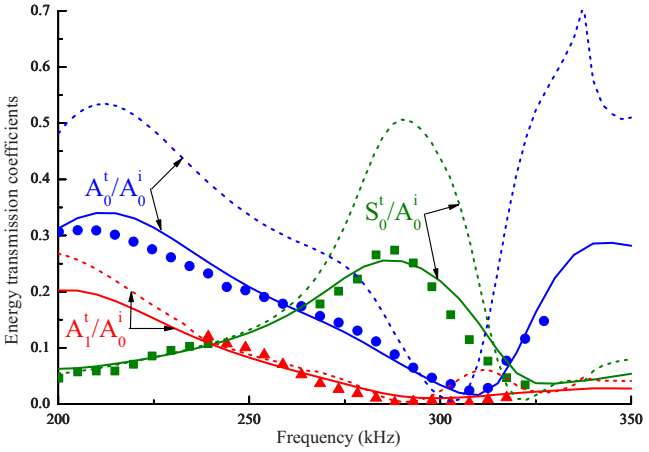


FIG. 10. (Color online) Transmitted energy ratios relative to A_0 incident; numerical results for elastic (—) and viscoelastic (—) cases and experimental data A_0 (●), S_0 (■), and A_1 (▲).

to the experimental energy flux ratios, as those predicted with the numerical model and orthogonality processing technique described in previous paragraphs.

A comparison between the numerical and experimental energy coefficients is shown in Fig. 10. Good agreement is obtained over the frequency range of investigation. Both the order of magnitude and the positions of peaks and minima are well confirmed by the experiments. This specific position is influenced by the material damping, as can be seen in Fig. 8(a), and the experiment proves the accuracy of the model. Concerning the A_1 mode below 250 kHz, its attenuation increases considerably as the frequency approaches the cutoff value, thus making it very difficult to be detected. Some missing experimental data for the S_0 mode between 240 and 260 kHz are very likely to be due to the strong dispersive behavior of this mode in this frequency range.

V. CONCLUSIONS

The influence of material viscoelasticity was investigated in a multimodal scattering problem, involving five scattered modes. The difficulty of modal identification in the scattered fields has been overcome by a hybrid method: FE steady state analysis of the structure at each frequency followed by a projection of the so-predicted fields on a modal basis using an orthogonality formula. The modal basis is based on the through-thickness displacement and stress fields obtained with a separate SAFE approach.

Two important effects of the material viscoelasticity on the scattered modes have been emphasized in this research work. The first one is a significant change in the amplitudes of the modes reflected by and transmitted past the defect. The second one is a slight shift of the frequencies at which minima/maxima of the energy fluxes occur in the scattered fields. These effects of the material viscoelasticity on the scattered modes have been shown to be more significant for defects of large extent in the direction of the waveguide. The numerical predictions of the modal energy fluxes transmitted by a 4 mm wide notch (the order of magnitude of the incident mode wavelength) have been confirmed by experimental measurements made with air-coupled ultrasonic transducers, the incident wave being the fundamental A_0 mode.

The results obtained in this study demonstrate that the material viscoelasticity must be used to solve guided wave scattering problems especially at certain frequency ranges where it has a strong influence on dispersion curves and modal shapes. This influence is especially important for inverse problems, which consist in characterizing defects.

ACKNOWLEDGMENTS

This work has been supported by the French Ministry of Education Nationale, de l'Enseignement Supérieur et de la Recherche through a PAST grant for collaborations with East European countries and also by the CNCSIS Romanian Research grant [Contract No. GR18 (123)/29.05.2007].

[1]D. Chimenti, "Guided waves in plates and their use in materials characterization," Appl. Mech. Rev. **50**, 247–284 (1997).
[2]M. Castaings, B. Hosten, and T. Kundu, "Inversion of ultrasonic, plane-wave transmission data in composite plates to infer viscoelastic material properties," NDT & E Int. **33**, 377–392 (2000).
[3]C.W. Chan and P. Cawley, "Lamb waves in highly attenuative plastic plates," J. Acoust. Soc. Am. **104**, 874–881 (1998).
[4]M. Castaings, C. Bacon, and B. Hosten, and M. V. Predoi, "Predoi, Finite element predictions for the dynamic response of thermo-viscoelastic material structures," J. Acoust. Soc. Am. **115**, 1125–1133 (2004).
[5]L. Moreau, M. Castaings, B. Hosten, and M. V. Predoi, "An orthogonality relation-based technique for post-processing finite element predictions of waves scattering in solid waveguides," J. Acoust. Soc. Am. **120**, 611–620 (2006).
[6]L. Duquenne, E. Moulin, J. Assaad, and S. Grondel, "Transient modeling of Lamb waves generated in viscoelastic materials by surface bonded piezoelectric transducers," J. Acoust. Soc. Am. **116**, 133–141 (2004).
[7]G. Shkerdin and C. Glorieux, "Lamb mode conversion in an absorptive bi-layer with a delamination," J. Acoust. Soc. Am. **118**, 2253–2264 (2005).
[8]J. M. Galán and R. Abascal, "Remote characterization of defects in plates with viscoelastic coatings using guided waves," Ultrasonics **42**, 877–882 (2004).
[9]B. Hosten, L. Moreau, and M. Castaings, "Reflection and transmission of guided waves reflected by defects in viscoelastic material plates," J. Acoust. Soc. Am. **121**, 3409–3417 (2007).
[10]J. M. Galan and R. Abascal, "Numerical simulation of Lamb wave scattering in semi-infinite plates," Int. J. Numer. Methods Eng. **53**, 1145–1173 (2002).
[11]J. M. Galan and R. Abascal, "Lamb mode conversion at edges. A hybrid boundary element–finite-element solution," J. Acoust. Soc. Am. **117**, 1777–1784 (2005).
[12]M. V. Predoi, M. Castaings, B. Hosten, and C. Bacon, "Wave propagation along transversely periodic structures," J. Acoust. Soc. Am. **121**, 1935–1944 (2007).
[13]S. Datta, A. Shah, R. Bratton, and T. Chakraborty, "Wave propagation in laminated composite plates," J. Acoust. Soc. Am. **83**, 2020–2026 (1988).
[14]P. J. Shorter, "Wave propagation and damping in linear viscoelastic laminates," J. Acoust. Soc. Am. **115**, 1917–1925 (2004).
[15]S. I. Rokhlin, T. K. Bolland, and A. Adler, "Splitting of domain of angles for incident wave vectors in elastic anisotropic media," J. Appl. Phys. **59**, 3672-3677 (1986).
[16]M. J. S. Lowe, "Matrix techniques for modeling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542 (1995).
[17]B. Hosten and M. Castaings, "Surface impedance matrices to model the propagation in multilayered media," Ultrasonics **41**, 501–507 (2003).
[18]COMSOL, User's Guide and Introduction, Version 3.2, COMSOL AB, 2005, http://www.comsol.com/
[19]A. Auld, *Acoustic Fields and Waves in Solids* (Wiley, New York, 1973).
[20]G. Shkerdin and C. Glorieux, "Lamb mode conversion in a plate with a delamination," J. Acoust. Soc. Am. **116**, 2089–2100 (2004).
[21]F. Simonetti and M. J. S. Lowe, "On the meaning of Lamb mode nonpropagating branches," J. Acoust. Soc. Am. **118**, 186–192 (2005).

Scattering of the fundamental shear horizontal guided wave by a part-thickness crack in an isotropic plate

P. Rajagopal and M. J. S. Lowe
Department of Mechanical Engineering, Imperial College London, SW7 2AZ, United Kingdom

(Received 26 November 2007; revised 16 July 2008; accepted 18 August 2008)

The interaction of the fundamental shear horizontal (SH0) guided mode with part-thickness cracks in an isotropic plate is studied as an extension within the context and general framework of previous work ["Short range scattering of the fundamental shear horizontal guided wave mode normally incident at a through thickness crack in an isotropic plate," J. Acoust Soc. Am. **122**, 1527–1538 (2007); "Angular influence on scattering when the fundamental shear horizontal guided wave mode is incident at a through-thickness crack in an isotropic plate," J. Acoust. Soc. Am. **124**, 2021–2030 (2008)] by the authors with through-cracks. The symmetric incidence case where the principal direction of the incident beam bisects the crack face at 90° is studied using finite element simulations validated by experiments and analysis, and conclusions are inferred for general incidence angles using insights obtained with the through-thickness studies. The influence of the crack length and the monitoring distance on the specular reflection is first examined, followed by a study of the angular profile of the reflected field. With each crack length considered, the crack depth and operating frequencies are varied. For all crack depths studied, the trend of the results is identical to that for the corresponding through-thickness case and the values differ only by a frequency dependent scale factor. Theoretical analysis is used to understand the physical basis for such behavior and estimates are suggested for the scale factor—exact for the high-frequency scattering regime and empirical for the medium- and low-frequency regimes. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2982410]

PACS number(s): 43.35.Zc, 43.20.Mv [YHB] Pages: 2895–2904

I. INTRODUCTION

The authors have been studying the interaction of cylindrical crested waves of the fundamental shear horizontal (SH0) guided mode with finite cracks in an isotropic plate, in the context of array imaging using ultrasonic guided waves.[1,2] The background to this research and studies of SH0 wave scattering by through-thickness cracks were presented in previous work[3,4] by the authors. The interaction was studied in terms of both the reflection from the crack face and the diffraction at the crack tips or edges. Finite element (FE) analysis was used to obtain results that were verified by experiments and comparison with those from similar problems in bulk elastic wave scattering literature. Figure 5 in Ref. 3 and Fig. 6 in Ref. 4 showing results from FE simulation of the interaction of SH0 waves with through-cracks at normal and oblique incidence angles, respectively, are illustrative of the configuration previously studied. A low-frequency point force excitation vibrating parallel to the plate surface (an "in-plane" force) as well as the crack length simulates the effect of an individual element of a transducer array. This generates circular waves of the SH0 mode principally perpendicular to the excitation direction and those of the fundamental symmetric Lamb mode (S0) principally parallel to it. Both of these modes have a cosine variation of amplitude away from their principal directions—this allows only a weak interaction of the S0 mode with the crack. The SH0 mode reflects from the crack face and is diffracted by its edges as well as undergoing conversion to the S0 mode.

These scattering studies with through-thickness cracks provided valuable insights into the physical principles involved and helped identify the minimum number of independent parameters required. The first studies on the spatial behavior of the scattering[3] revealed the influence of diffraction phenomena in the short range leading to focusing of energy by the cracks in the backscatter direction. The diffracted field itself consists of contributions due to a direct diffraction at the crack edges and reverberations of Rayleigh-like[5,6] waves along the crack length. A simple diffraction model assuming uniform shear sources on the insonified face is found to adequately represent cracks up to moderate lengths and thus provides an easy means of estimating the far field of the waves. Studies of the angular profile of the reflection[4] suggested that it is mainly concentrated in a narrow band around the specular direction, approaching that from a straight edge for long cracks. Specular reflection itself is strongest in the case of normal or symmetric incidence (when the central "ray" of the incident beam bisects the crack face at 90°) and varies rapidly at other angles, dropping down to zero close to the first critical angle of the material. If the crack and signal length are appropriate to allow isolation of the primary diffraction, its angular profile around the crack edge can be estimated from known solutions for bulk shear vertical (SV) wave canonical scattering problems.[7–10]

Here we further extend our investigation to consider part-through instead of through-thickness cracks. Although we only study the normal or symmetric incidence case and the reflection behavior, based on the relationship of the re-

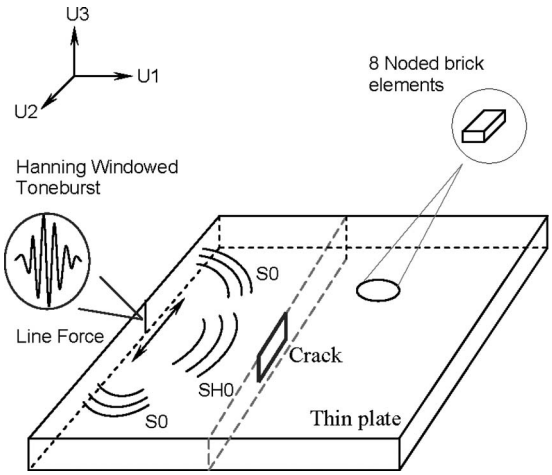


FIG. 1. Details of the configuration studied and those for the FE simulations (Plate thickness shown exaggerated).

sults with those for through-cracks we draw general conclusions for the scattering of the SH0 mode. In doing so, we will again adopt the method in Refs. 3 and 4: We first examine the influence of the crack length and the monitoring distance on the specular reflection and then study the angular profile of the diffuse reflection. Here the term "specular" refers to the direction of mirrorlike reflection as given by Snell's law and "diffuse" to all the directions in which the waves can be reflected in the backscatter direction. To facilitate comparison with earlier through-thickness crack results, the same excitation and monitoring distances from the crack as used for similar studies in Refs. 3 and 4 are maintained as far as possible. For each crack length, different crack depths and operating frequencies are considered, to gain an understanding of the reflection behavior. Studies on the relation of the scattering from part-thickness cracks and notches have shown[11] that notches of extremely small axial extent or width ($<0.01\%$ of the wavelength) are necessary to approximate the behavior of cracks. Instead of attempting to create such small-width notches, separate FE studies with finite-width notches were performed for experimental validation of the modeling.

The sections of this paper are organized as follows. In Sec. II we describe the configuration studied and the general procedure for FE simulations and experiments. Section III deals with specific studies, taking up in order, the influence of the crack length and measurement location on the specular reflection, the angular profile of the diffuse reflected field, and finally the validation of FE modeling with experiments. In Sec. IV, we analyze the results and seek an empirical formula for the scattering of the SH0 mode by part-thickness cracks and draw conclusions in Sec. V.

II. METHODS

A. Configuration studied

Figure 1 shows the details of the configuration studied and those for the FE simulations described in the next part of this section. A finite length ideal (planar, zero-width) surface-breaking part-thickness crack is located in a thin isotropic plate. A line excitation consisting of in-plane point forces vibrating in the U2 direction applied uniformly through the plate thickness generates[12] incident circular-crested waves of the SH0 mode principally along U1 and those of the S0 mode principally along U2. The operating frequency-thickness values are always low enough so that only fundamental modes can be present even after scattering. Due to the through-thickness symmetry of the excitation, the fundamental antisymmetric Lamb mode (A0) is not generated at the source. The SH0 mode is generated such that it propagates principally along the crack's symmetric normal. The fact that the S0 mode travels much faster than the SH0 mode at these frequencies is used to time-gate its effects and although some A0 will be generated by mode conversion at the scattering, we make use of its modal properties so that it is not picked up at the monitored positions. We thus study in isolation the SH0 reflection from the crack face due solely to the SH0 mode incident at it.

We examine successively the influence of the crack length, depth, and measurement position on the reflection using FE simulations validated with experiments. We observed in Ref. 3 that SH0 mode scattering from through-thickness cracks, which are not too small (cracks of length ℓ such that $k_{SH0}\ell > 1$ where k_{SH0} is the wavenumber), can be described well by the Kirchhoff approximation. This means that the faces of the crack do not interact and reflection would be governed by the secondary sources placed on the insonified face. Thus the results at a frequency f_0 would be valid at another frequency f_1 provided the crack length and the other dimensions are scaled appropriately by a factor of f_1/f_0 by which the wavelength is scaled. However, such simple scaling will not work with part-thickness cracks since their faces interact in a more complex frequency-dependent manner and so the Kirchhoff approximation does not hold except at high frequencies: Therefore the influence of the operating frequency on each of these cases is also studied.

Mention must be made here of the analysis on the relationship between cracks and notches in Ref. 11, which brings out an important difference in the scattering behavior of a through-thickness notch from that of a part-thickness one. Only the insonified face of the through-thickness notch participates directly in the scattering process. However, with a part-thickness notch both faces contribute and the scattered packet consists of reflections from the first face followed by multiple reverberations across the "axial extent" or width of the notch. Because of this, the scattering from a part-thickness notch is strongly influenced by the operating frequency.

Thus a through-thickness crack can be reasonably well represented in its scattering by a through-thickness rectangular notch or slit with suitably small axial extent (width). However, a part-thickness notch cannot represent a part-through crack at all unless it has a width less than 0.01% of the operating wavelength. Since it is difficult to create such a small-width uniform notch, we decided to perform separate FE studies with notches for comparison with experiments. The configuration studied in this case remains the same, except that the crack is now replaced with a finite width notch. The dimensions of the notch and the operating frequency will all have a combined influence in this study.

B. General procedure for finite element simulations

The problem was simulated by FE analysis in the general three-dimensional domain implemented using the commercial package ABAQUS.[13] The model plate was assumed to be 2 mm thick and the material properties of aluminum were used, simply in order to match the experimental plate dimensions and properties. The mesh consisted of perfectly cubic eight noded linear continuum elements and four elements were used to represent the thickness. This allowed for around 16 elements per SH0 wavelength at the highest center frequencies used, thus respecting well the spatial discretization limit required for accurate modeling of wave propagation.[14] Cracks of vanishing width were created by disconnecting nodes on elements representing the adjacent faces. Notches of 0.5 mm width (equal to the size of an element) were created by removing columns of elements from the mesh. Absorbing layers with increasing damping (see Refs. 15–17 for details on the use of absorbing boundaries in FE analysis) were used around the edges of the plate in order to isolate the scattered SH0 waves.

The excitation consisted of a 5 cycle Hanning windowed toneburst centered at the required frequency and applied as a force in the U2 direction parallel to the crack face, at a single line formed by all nodes through the thickness. This generated the S0 and SH0 modes, with their principal directions parallel and perpendicular, respectively, to the applied force and ensures that no A0 mode is generated. Explicit time integration with a constant time step respecting the stability limit[14] simulated the propagation of the modes. The default condition on the faces of the disconnected elements representing the defect is that of zero stress. Thus the scattering from an ideal open crack or uniform rectangular notch was simulated.

As in Refs. 3 and 4, the reflection behavior was studied in terms of the frequency domain ratio of the beam-compensated reflected and incident signals, given as follows:

$$\text{Reflection ratio} = \frac{|R(\omega)|\sqrt{D+d_M}}{|I(\omega)|\sqrt{d_I}}, \tag{1}$$

where $R(\omega)$ and $I(\omega)$ are the frequency spectra of the resultant displacement obtained from the reflected and incident signals, D is the distance between the source and the center of the crack, d_M is the distance from the crack center to the point where the reflection is monitored, and d_I is the distance from the source to the point where the incident signal is monitored. (See, for instance, Fig. 4 for an illustration).

The reflection ratio for a fully through-thickness crack was obtained at 0.4 MHz mm and the results agreed very well with the previous plane-stress studies obtained at 0.1 MHz mm, thus assuring us of the integrity of the definition of the crack. As expected, the results for the through-thickness crack were identical for different frequencies, with the dimensions suitably scaled.

C. General procedure for experiments

The experimental setup was identical to the one described in detail in Refs. 3 and 4. A pointlike excitation was achieved by means of a plane wide-band piezoelectric shear transducer (Panametrics V301, 0.5 MHz center frequency) coupled to the plate through a small, thin brass disk. The source signal consisting of a 5 cycle Hanning windowed toneburst was generated by a wavemaker (Macro Design Ltd., UK) instrument and centered at the required frequency. The detection was achieved using a laser interferometer (Polytec OFV 2700, with dual differential fiber optic lines) measuring the displacement component parallel to the surface of the plate through small patches of reflective tape, which enhance the optical backscatter. The quality of each displacement measurement was improved by applying a bandpass filter and taking an average over 500 acquisitions.

Measurements were carried out on two 0.5 mm wide, 10 mm long notches of depths 1 and 1.5 mm, respectively, cut in 1250 mm square standard commercial purity aluminum plates of 2 mm thickness. The center frequency of excitation was varied in order to obtain different values of frequency-thickness and notch length in terms of the wavelength. The excitation and detection positions were also varied at each center frequency, so that they were located at the same number of wavelengths away from the notch as in the FE simulations.

III. SPECIFIC STUDIES AND RESULTS

A. Study of specular reflection

We first study the influence of the crack length and measurement position on the specular reflection, as the crack depth and the operating frequency are varied. Since this is a symmetric incidence problem, the direction of specular reflection lies along the symmetric normal to the crack face. In this direction the scattered waves of the A0 mode have displacements entirely in the U1-U3 plane, while those of the SH0 mode have displacements entirely in the U2 direction. Thus it was enough to monitor the U2 displacements on the surface of the model plate, in order to measure the reflected SH0 waves without contamination by the A0 mode. As mentioned earlier, the S0 mode travels much faster than the other two modes so that its effects can be time-gated out and thus SH0 mode scattering is studied exclusively.

1. Influence of crack length

A number of models with cracks of lengths up to 5 λ_{SH0} (SH0 wavelengths at the center frequency of the toneburst) were set up to understand the dependence of the SH0 reflection on the length of the crack. For each crack length, models with three different crack depths, 25%, 50%, and 75%, of the plate thickness were created. Further, in order to study the influence of the frequency-thickness, the simulations were performed with center frequencies of 200, 300, and 400 kHz for each crack depth. The excitation was located at a distance of 16 λ_{SH0}, while nodal displacements in the U2 direction were monitored at 5 λ_{SH0} from the crack and along its symmetric normal. The generated S0 mode has nearly zero displacement along this line, so only the incident and reflected SH0 signals are recorded at the monitored points.

The reflection ratio was calculated according to Eq. (1) for each center frequency value. Figures 2(a)–2(c) show it plotted with crack length (expressed in λ_{SH0}) for each of the

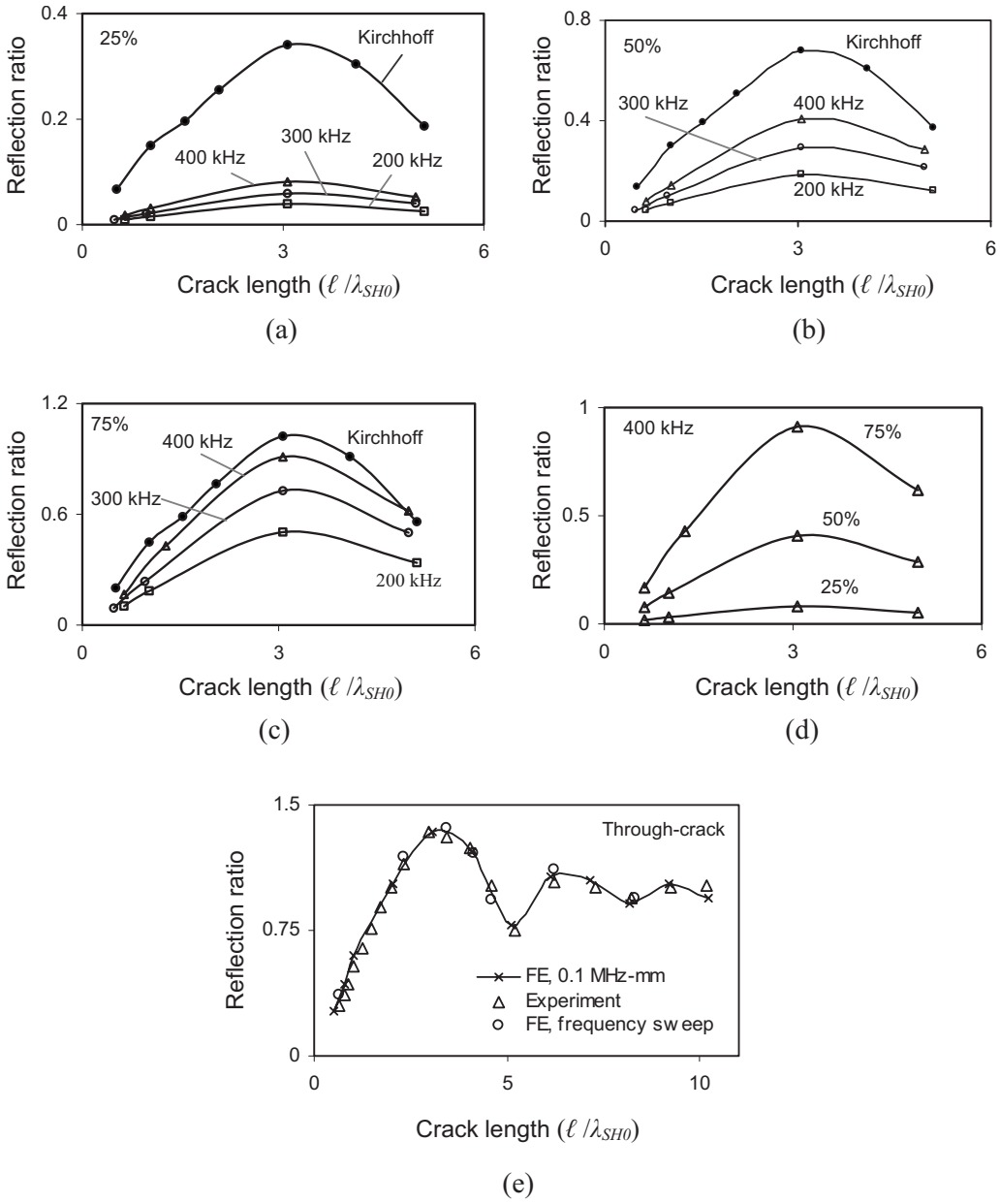


FIG. 2. [(a)–(c)] Reflection ratio plotted with crack length (expressed in λ_{SH0}) for 25%, 50%, and 75% deep cracks, respectively. Each plot shows the results at different incident frequencies; the lines with filled circles indicate the estimates using Kirchhoff approximation; (d) shows the results for all three crack depths for just the 400 kHz case; (e) reproduces Fig. 10 from Ref. 3.

three crack depths studied. Each plot shows the ratio at the three different frequency-thickness values and also the expected results if the Kirchhoff approximation was valid [obtained simply as 0.25, 0.5, and 0.75 times the results from the through-thickness crack case, reproduced in Fig. 2(e) from Ref. 3]. Figure 2(d) extracts from Fig. 2(a)–2(c) the results for all the three crack depths studied at a single frequency of 400 kHz.

The observed trend of the reflection ratio, rising almost linearly with crack length initially and then falling, agrees very well with that for through-thickness cracks where it later continues into an oscillatory regime. This trend is consistent with the understanding obtained from earlier studies: Diffraction effects are important in the short range, leading to focusing of energy in the backscatter direction. The actual reflection ratio values though fall short of the Kirchhoff prediction even for 75% deep cracks at low frequency-thickness, but the ratio rises sharply and can be seen to approach it asymptotically, as the frequency is increased. For shallow cracks 25% deep, the ratio is well below the approximate estimate even at 400 kHz and there is not much variation from that at 200 kHz. We will discuss this in some more detail with the analysis in Sec. IV.

2. Influence of distance of measurement

The source was fixed at a distance of 32 λ_{SH0} from the crack face, while the reflection was monitored at distances of 3, 4, 5, 10, 15, 20 and 25 λ_{SH0} from it. Results were obtained for two crack lengths, 1 and 5 λ_{SH0}, representing small and large cracks, respectively. For each crack length two crack depths, 25% and 75%, of the plate thickness, representing

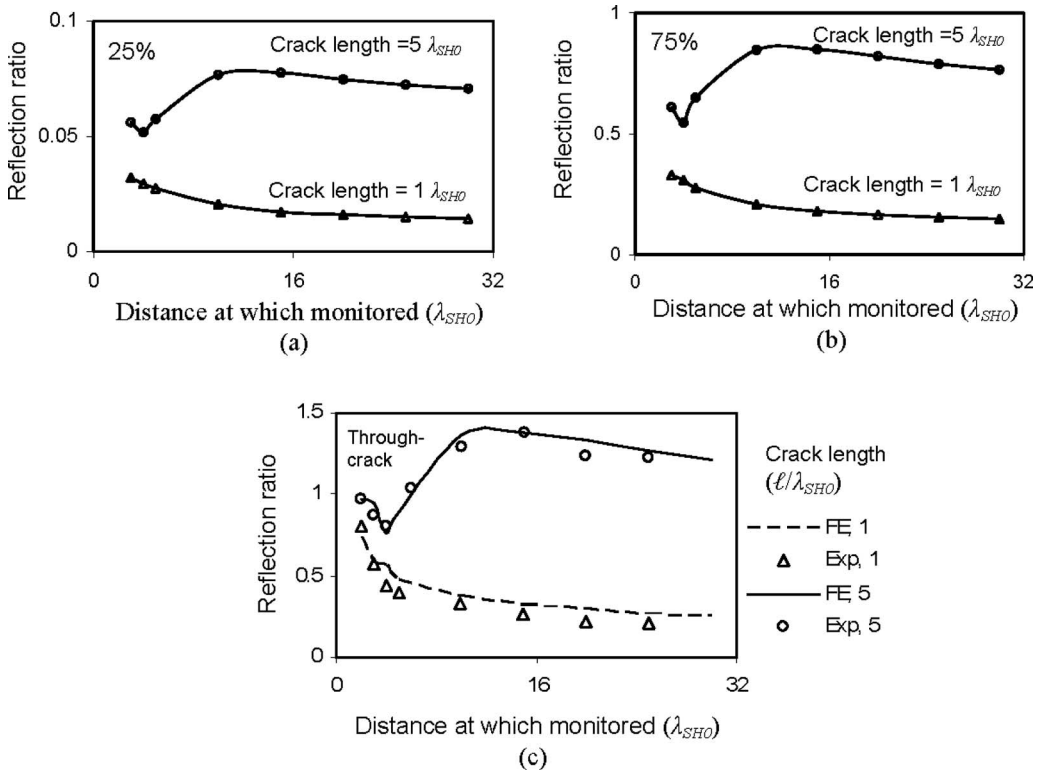


FIG. 3. Reflection ratio plotted with the distance at which the reflected signal is monitored (expressed in λ_{SH0}) for two crack lengths 1 λ_{SH0} and 5 λ_{SH0}; [(a) and (b)] results for two crack depths, 25% and 75%, of the plate thickness; (c) results for through-thickness crack cases of the same lengths, obtained from Ref. 3.

shallow and deep cracks, were studied, and simulations were performed at a center frequency of 400 kHz spanning the low- to high-frequency regimes.

The reflection ratio was calculated at this center frequency, again using Eq. (1). Figures 3(a) and 3(b) show the reflection ratio plotted with the distance at which the reflected signal is monitored (expressed in λ_{SH0}) for the two crack length cases and each of the two crack depths studied. Figure 3(c) reproduces the results for 1 and 5 λ_{SH0} long through-thickness cracks from Ref. 3. We notice that at both crack depths the trend is identical with that for the through-thickness crack case of the same length except for a scale factor associated with the part-through cracks. After an initial fluctuation, the reflection tends to settle down at increasing distances from the crack. For through-thickness cracks, this behavior could be explained by a simple model assuming the insonified face to behave like a uniform ideal shear transducer. The present study shows that such an explanation holds even for part-thickness cracks and the results differ only in their magnitude by a frequency-dependent factor. We will examine the nature of this factor in Sec. IV.

B. Angular profile of diffuse reflection

Next we examine the reflection at various angles θ_R for an incidence angle $\theta_I=0°$ (the angles are defined counterclockwise from the crack's symmetric normal). At these diffuse angles the reflected A0 mode will have displacements in all the three directions at the surface of the plate, but its mode shape is such that the in-plane components vanish entirely at the midplane. Therefore for this set of studies, the SH0 mode reflection is isolated by taking all measurements at the midplane of the model plate.

Figure 4 illustrates the geometry used for this purpose. The excitation was located at a distance $D=32$ λ_{SH0} (SH0 wavelengths at the center frequency of the toneburst) away from the crack and such that the principal axis (PA) of the generated SH0 waves coincides with SN the symmetric normal to the crack. The incident signal I was monitored at a distance d_I from the source, along PA. Reflected SH0 waves were monitored along one-half of a vertical line intersecting SN at a point S and a distance $d_{M0}=12$ λ_{SH0} from the crack. Starting from a point along this line at a distance of 12.5 λ_{SH0} from S, nodal displacements R_{U2} in the U2 direction

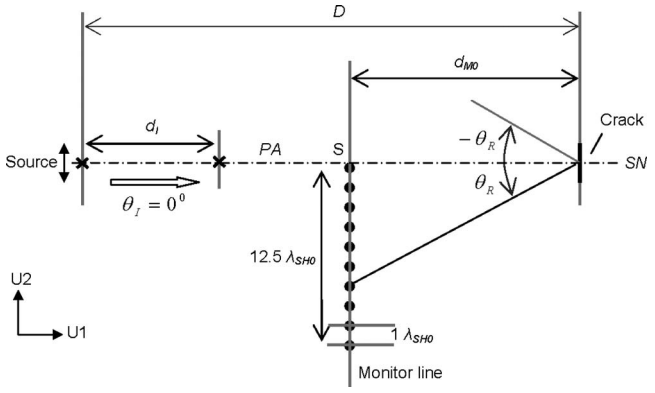


FIG. 4. Geometry for the study of reflection at angles θ_R for incidence angle $\theta_1=0°$.

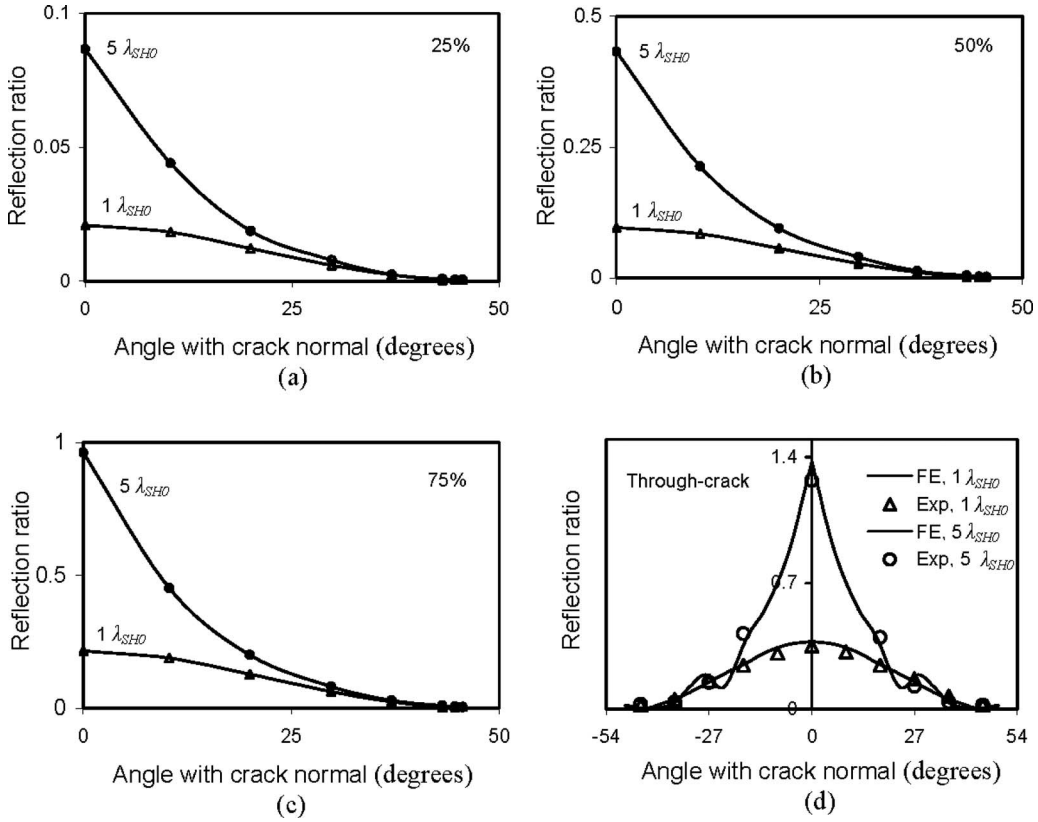


FIG. 5. Reflection ratio plotted with θ_R: The results are shown for two crack lengths 1 λ_{SH0} and 5 λ_{SH0} studied for each of three crack depths cases (a) 25%, (b) 50%, and (c) 75% of the plate thickness; (d) results for through-cracks of the same lengths from Ref. 4.

were monitored at intervals of around 1 λ_{SH0} all the way down to S. This yielded monitoring angles in the range $\theta_R = 0° - 45°$ in intervals of 6°–10°. The resultant displacement R is related to R_{U2} by $R\cos(\theta_R) = R_{U2}$ so the resultant displacement values were obtained from the U2 values for each monitored point as $R = R_{U2}/\cos(\theta_R)$. The values in the range from −45° down to 0° were assumed to be identical with these due to the symmetry in the configuration. The reflection ratio was then calculated using Eq. (1) with $d_M = d_{M0}/\cos(\theta_R)$. Cracks of two lengths 1 and 5 λ_{SH0}, and three different depths 25%, 50%, and 75% of the plate thickness for each crack length were studied, and simulations were performed at a center frequency of 400 kHz.

Figures 5(a)–5(c) show the reflection ratio plotted with the angle at which the reflection is monitored for the two crack length cases and each of three crack depths studied, while Fig. 5(d) presents the corresponding results for through-thickness cracks from Ref. 4. We again observe excellent agreement of the angular profile of the reflected field with that for the through-thickness crack case for all the three crack depths. The reflected energy is concentrated in a narrow beam around the specular direction and this focusing increases with the length of the crack, just as an ideal transducer would tend to generate a more collimated beam at larger diameters.

C. Experimental validation of FE modeling

The reflection ratio–crack length study in Sec. III A 1 was repeated with the crack replaced with a notch, for the validation studies. 10 mm long, 0.5 mm wide notches at two depths, 50% and 75%, of the plate thickness were created in the FE models, matching the details of the notches cut in the experimental plates. In both simulations and experiments, the same procedure was followed regarding the excitation and monitoring of scattered signals and the frequency sweep. SH0 waves were excited from a distance of 16 λ_{SH0} from the notch face such that a symmetric incidence problem is realized. Nodal displacements in the U2 direction were monitored at 5 λ_{SH0} from the notch, along its symmetric normal. As in the crack case, the incident and reflected SH0 signals are expected almost exclusively at the monitored points. Starting from 200 kHz, the operating frequency was swept in steps of 50 up to 400 kHz so that different notch lengths with respect to λ_{SH0} were obtained. The reflection ratio was again calculated using Eq. (1), and Fig. 6 shows the results obtained at different frequency-thickness (and hence different ℓ/λ_{SH0}) values for the two notch depths, from FE simulations and experiments. We find excellent agreement between the measurements and predictions.

IV. ANALYSIS

The studies in the previous sections show that the reflection of the SH0 mode from part-thickness cracks is related to that from through-thickness cracks of the same length through a frequency-dependent scale factor. This factor as it were seems to account for the change in the crack depth. Through the following parts of the present section, we seek to systematically understand the physical basis for and the

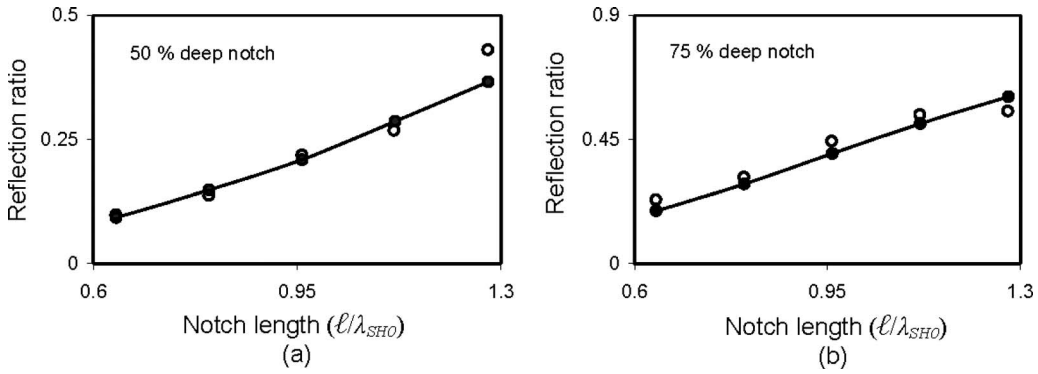


FIG. 6. Reflection results from FE simulations (filled circles) and experiments (empty circles) obtained using a frequency sweep (and hence different ℓ/λ_{SH0}) for two notch depth cases: (a) 50% and (b) 75% of the plate thickness.

conditions underpinning such behavior. Continuing with the analytical approach of Ref. 3, in Sec. IV A we derive a formula for the beam-compensated reflection ratio for our problem. Consideration of the relationship between the reflection ratio for through- and part-thickness cracks in Sec. IV B using insights from Refs. 3 and 4 reveals the conditions when a scale-factor-like behavior manifests. Finally, Sec. IV C helps us understand the nature of the scale factor by arriving at and testing the limits of estimates for it in the different frequency regimes.

A. Formula for reflection ratio

Following the analysis in Ref. 3, the scattered displacement field $u_k^{sc}(\vec{p})$ in the $\hat{e}_k$ direction at a position $\vec{p}$ in the medium due to a crack of length ℓ, depth b, and outward normal n_j is written as a superposition of elemental sources placed on its insonified face s^+ as[18,19]

$$u_k^{sc}(\vec{p},\omega) = \int_{s^+} \Delta u_i(\vec{s},\omega)[\Sigma_{ij;k}^0(\vec{p}|\vec{s},\omega)n_j]dS, \tag{2}$$

where $\Delta u_i(\vec{s})$ is the displacement jump across the crack faces, also known as the crack opening displacement or "COD", $\Sigma_{ij;k}^0(\vec{p}|\vec{s},\omega)n_j$ is the traction Green's function giving the traction force in the $\hat{e}_k$ direction at a position $\vec{p}$ due to a point force applied in the $\hat{e}_i$ direction at the position $\vec{s}$ in *the free medium in the absence of the crack*, dS is an area element, and the integral is on the insonified face s^+ of the crack. As usual, the frequency term $e^{j\omega t}$ is suppressed here and throughout this presentation. Here we must clarify that we use the vertical bar symbol "|" in the sense of "given" instead of the "forward slash" symbol "/" (as, e.g., in $\vec{p}/\vec{s}$) used in Ref. 3 to denote the same, to avoid confusion with the latter's usual meaning as "divided by."

In the following, we will adapt Eq. (2) to our problem. A harmonic point excitation vibrating parallel to the crack face and uniform through the thickness of the thin plate is assumed to be located at a distance D from the crack. We introduce cylindrical coordinates as shown in Fig. 7, with the origin at the excitation point and the positive z axis pointing into the plate thickness. Thus the excitation consists of a line force vibrating along $\theta=\pi/2$ and uniform through the plate thickness, while the crack is located along the line $r\cos(\theta)=D$ with normals $\hat{n}=\cos(\phi)\hat{e}_r+\sin(\phi)\hat{e}_\theta+z\hat{e}_z$.

Introducing $d\ell=rd\theta$ as a length element and as in Ref. 3, taking $\cos(\phi)\approx 1$, the scattered displacement due to the incident SH0 mode having displacements only in the u_θ direction can then be given as

$$u_k^{sc}(\vec{p},\omega) = \int\int_{s^+} \Delta u_\theta(\vec{s},\omega)\Sigma_{r\theta;k}^0(\vec{p}|\vec{s},\omega)d\ell dz. \tag{3}$$

At the low frequencies considered, displacements and stresses will be uniform in the z or the depth direction, so we can take $\Sigma_{r\theta;k}^0(\vec{p}|\vec{s},\omega)$ as depth invariant; assuming further that the displacement jump $\Delta u_\theta(\vec{s},\omega)$ does not vary in the length direction of the crack and changing the order of integration yields

$$u_k^{sc}(\vec{p},\omega) = \int_0^b \Delta u_\theta(\vec{s},\omega)dz \int_\ell \Sigma_{r\theta;k}^0(\vec{p}|\vec{s},\omega)d\ell. \tag{4}$$

The reflection ratio according to Eq. (1) at a required position $\vec{p}$ in the medium for a crack of depth b and length ℓ would then be

$$R(b,\ell) = \frac{u_\theta^{sc}(\vec{p},\omega)\sqrt{D+d_M}}{u_\theta^{inc}(\vec{s},\omega)\sqrt{D}} = \sqrt{\frac{D+d_M}{D}}\frac{\int_0^b \Delta u_\theta(\vec{s},\omega)dz}{u_\theta^{inc}(\vec{s},\omega)}\int_\ell \Sigma_{r\theta;\theta}^0(\vec{p}|\vec{s},\omega)d\ell, \tag{5}$$

where $u_\theta^{inc}(\vec{s},\omega)$ is the incident SH0 displacement on the insonified crack face.

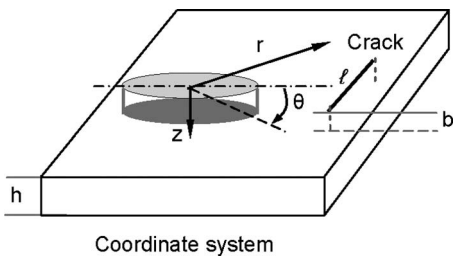


FIG. 7. Cylindrical coordinates with origin at the excitation point and the positive z axis pointing into the plate thickness: A line force vibrating along $\theta=\pi/2$ and uniform through the plate thickness provides the excitation and the crack is located along the line $r\cos(\theta)=D$ with normals $\hat{n}=\cos(\phi)\hat{e}_r+\sin(\phi)\hat{e}_\theta+z\hat{e}_z$.

B. Formula for the scale factor

If the Kirchhoff approximation in the depth direction is valid, then $\Delta u_\theta(\vec{s},\omega)=u_\theta(\vec{s}\in s^+,\omega)=2u_\theta^{\text{inc}}(\vec{s},\omega)$ so that the first integral in Eq. (4) is easily evaluated: $\int_0^b \Delta u_\theta(\vec{s},\omega)dz = 2u_\theta^{\text{inc}}(\vec{s},\omega)b$. The reflection ratio would then simply be

$$R(b,\ell)=\sqrt{\frac{D+d_M}{D}}2b\int_\ell \Sigma^0_{r\theta;\theta}(\vec{p}|\vec{s},\omega)d\ell\,. \tag{6}$$

As observed earlier this approximation holds for through-thickness cracks that are not too small (i.e., of length ℓ such that $k_{\text{SH0}}\ell>1$) and we can take the reflection ratio $R(b,\ell)$ for such a case to be given by Eq. (6) with reasonable accuracy.

Thus for part-thickness cracks that can exhibit a Kirchhoff-like behavior, the reflection ratio for a crack of a given depth b and length ℓ can be obtained from the value for a through-thickness crack of the same length simply as

$$R(b,\ell)=\alpha R(h,\ell) \quad \text{with } \alpha=b/h, \tag{7}$$

where h is the thickness of the thin plate.

In general though, the reflection ratio for a part-thickness crack would be related to that for the through-thickness case as

$$R(b,\ell)=k(\alpha,\omega)R(h,\ell), \tag{8a}$$

where

$$k(\alpha,\omega)=\frac{\int_0^b \Delta u_\theta(\vec{s},\omega)dz}{(2h)u_\theta^{\text{inc}}(\vec{s},\omega)}. \tag{8b}$$

Engineering intuition would tell us that we should be able to obtain the reflection from a part-thickness crack by knowing that from a through-crack of the same length and accounting for the difference in the crack depth. However, such thinking implicitly takes the effects of the length and the depth of the crack on the scattering process to be uncoupled. The assumptions made in arriving at Eqs. (8a) and (8b) uncover the conditions in which this uncoupling can happen: The frequency must be low enough such that the stresses are constant through the plate thickness, and the crack must be small compared to the incident wavefront, so that the COD can reasonably be assumed to be constant along the crack length. Equation (8b) reveals that the frequency dependence of the scale factor $k(\alpha,\omega)$ comes from that of the amplitude of the incident wave and how the crack opens in response. Also we see that the difference in the reflection from part- and through-thickness cracks is due to the difference in the way they respond to the incident wave.

C. Estimates

Specific knowledge of how the crack opens is necessary to evaluate the factor $k(\alpha,\omega)$ by which the part- and through-thickness crack cases are related. However, here approximate estimates for the different scattering regimes are suggested. For the high-frequency regime we can deduce from Eq. (7) that $k(\alpha,\omega)=\alpha$. Here we must mention that we use the terms "high," "medium" and "low" only as relative indicators within the overall context of the low frequency-thickness values of interest to this paper. From past studies on the scattering of fundamental guided waves by long cracks,[20,21] at low frequencies $\Delta u_\theta(s,\omega)$ would appear to be of the order of $(b/h)^2$ or α^2: Therefore, we can take $k(\alpha,\omega)\approx\alpha^3$. This seems valid for the low-frequency scattering of SH0 waves by long cracks, as can be inferred from the results of Demma *et al.*[22] Finally, through an examination of the results presented in Sec. III, a suggested empirical first estimate for medium frequencies is $k(\alpha,\omega)\approx\alpha^2$.

TABLE I. Cracks of different depths b studied in Sec. III A 1 are classified into the (relatively within the overall low-frequency context) low- (L), medium- (M), or high- (H) frequency regimes based on whether the product $k_{\text{SH0}}b$ is less than, approximately equal to, or greater than unity.

Crach depth→ ↓ Frequency	25% 0.5 mm	50% 1 mm	75% 1.5 mm
200 kHz $k_{\text{SH0}}\approx 0.4$ mm^{-1}	L	L	L
300 kHz $k_{\text{SH0}}\approx 0.6$ mm^{-1}	L	L	M
400 kHz $k_{\text{SH0}}\approx 0.8$ mm^{-1}	L	M	M/H

We will next examine the limits of these estimates, taking as an example the results from Sec. III A 1. Table I classifies cracks of different depths b considered in that study into the low-, medium-, or high-frequency regimes based on whether the product $k_{\text{SH0}}b$ is less than, approximately equal to, or greater than unity. This helps us identify the estimates for $k(\alpha,\omega)$ in the frequency regime appropriate for each crack depth. Equation (8a) is then invoked and the reflection ratio for the different part-thickness cracks is calculated from the results for through-thickness cracks of the same length taken from Ref. 3. The estimated values of the reflection ratio so obtained are compared with the actual FE results of Sec. III A 1 in Figs. 8(a)–8(c) for different crack depths and frequency-thickness values. Figure 8(d) shows the same comparison but with the reflection ratio plotted with the crack depth instead of crack length, for a 3 λ_{SH0} long crack. We find very good agreement for the 75% crack depth case where the three frequencies studied nearly fall into the three different scattering regimes, whereas for the other two crack depths, the estimates are not that accurate. For the 25% deep crack, Table I would suggest that all the three frequencies studied lie within the low-frequency regime, but the actual FE results lie somewhere between the low- and medium-frequency estimates. For the 50% deep crack, even though the full FE results lie around the low- to medium-frequency regimes, the estimates cannot distinguish between 300 and 400 kHz. Arguably the low-frequency estimates suggested here are very rough, but the results show that often we can obtain a fair hint of the order of the reflection amplitude. With better approximations for $k(\alpha,\omega)$ and knowledge of the crack length and frequency used, this approach can provide an alternative prognosis for the crack depth.

V. DISCUSSION AND CONCLUSIONS

This paper studied the interaction of low-frequency SH0 waves with part-thickness cracks for the symmetric inci-

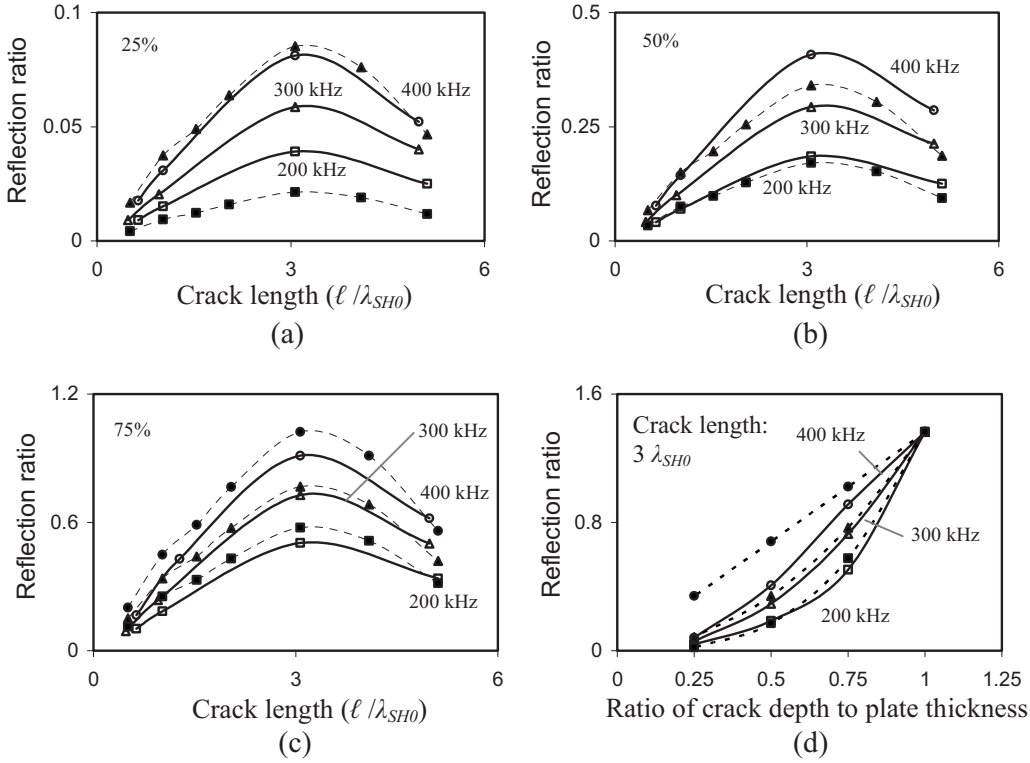


FIG. 8. Estimated values of reflection ratio obtained using the approach in Sec. IV compared with the actual FE results of Sec. III A 1 for different incident frequencies at each of the crack depths, (a) 25%, (b) 50%, and (c) 75% of the plate thickness: The filled square, triangle, and circle denote the estimates in the low-, medium -, and high-frequency regimes, respectively; (d) shows the same information but the reflection ratio is plotted with the crack depth instead of length, for the case of a 3 λ_{SH0} long crack.

dence problem. The interaction was studied in terms of the influence on the specular reflection, of the crack length, and the spatial and angular measurement positions at different crack depths and operating frequencies. The FE modeling was validated by comparison with experiments for the case of scattering by a finite part-depth rectangular notch. The results from all the studies are similar to those obtained for the corresponding through-thickness case with the trend being identical and actual values differing only by a frequency-dependent scale factor. Theoretical analysis is used to reveal the physical basis for such behavior and helps us obtain estimates for the scale factor, exactly derived for the high-frequency scattering regime and empirical for the medium- and low-frequency regimes.

From the perspective of guided wave imaging, the uncoupling of the effects of the length and depth of cracks as suggested by these results means that the understanding obtained for scattering of the SH0 mode from through-thickness cracks is often sufficient to deal with the part-thickness case as well. Although only the symmetric incidence problem and the reflected field were studied here, the generality of the results helps us infer that the reflection at oblique or nonsymmetric incidence as well as the diffraction from part-cracks would be related to that for the through-thickness case through an appropriate factor. Estimates of the scaling and thus of the scattered amplitudes at different scattering regimes will help when studying the influence of the crack dimensions on the image intensity. Finally, the influence of the width of a finite notch on the scattering as highlighted here again brings up the importance of the difference between notches and cracks and the frequency-related limit when they would be equivalent.

[1]J. Davies, F. Simonetti, M. J. S. Lowe, and P. Cawley, "Review of synthetically focused guided wave imaging techniques with application to defect sizing," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2006), Vol. 25, pp. 142–149.
[2]J. Davies and P. Cawley, "The application of synthetically focused imaging techniques for high resolution guided wave pipe inspection," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2007), Vol. 26, pp. 681–688.
[3]P. Rajagopal and M. J. S. Lowe, "Short range scattering of the fundamental shear horizontal guided wave mode normally incident at a through-thickness crack in an isotropic plate," J. Acoust. Soc. Am. **122**, 1527–1538 (2007).
[4]P. Rajagopal and M. J. S. Lowe, "Angular influence on scattering when the fundamental shear horizontal guided wave mode is incident at a through-thickness crack in an isotropic plate," J. Acoust. Soc. Am. **124**, 2021–2030 (2008).
[5]B. K. Sinha, "Some remarks on propagation characteristics of ridge guides for acoustic surface waves at low frequencies," J. Acoust. Soc. Am. **56**, 16–18 (1974).
[6]A. A. Oliner, "Waveguides for acoustic surface waves: A review," Proc. IEEE **64**, 615–627 (1976).
[7]R. K. Chapman, "Ultrasonic reflection from smooth flat cracks: Exact solution for the semi-infinite crack," CEGB Report No. NW/SSD/RR/14/81, N.D.T Applications Centre, Scientific Services Department, Central Electricity Board, UK, North Western Region, 1981.
[8]J. A. Ogilvy and J. A. G. Temple, "Diffraction of elastic wave by cracks: Application to time-of-flight inspection," Ultrasonics **21**, 259–269 (1983).
[9]J. D. Achenbach, A. K. Gautesen, and H. Mcmaken, *Ray Methods for Waves in Elastic Solids* (Pitman Books Ltd., London, 1982).
[10]C. B. Scruby, K. R. Jones, and L. Antoniazzi, "Diffraction of elastic waves by defects in plates: Calculated arrival strengths for point force and thermoelastic sources of ultrasound," J. Nondestruct. Eval. **5**, 145–156 (1986).

[11]A. Demma, P. Cawley, and M. Lowe, "Scattering of the fundamental shear horizontal mode from steps and notches in plates," J. Acoust. Soc. Am. **113**, 1880–1891 (2003).
[12]P. Wilcox, "Modeling the excitation of Lamb and SH waves by point and line sources," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2004), Vol. 23, pp. 206–213.
[13]ABAQUS V6.6 Analysis User's Manual. For information on ABAQUS FE package see http://simulia.com/products/abaqus_fea.html (last viewed July, 2008).
[14]D. N. Alleyne and P. Cawley, "The interaction of Lamb waves with defects," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 381–397 (1992).
[15]M. Drozdz, M. J. S. Lowe, E. Skelton, and R. Craster, "Modeling bulk and guided wave propagation in unbounded elastic media using absorbing layers in commercial FE packages," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2007), Vol. 26, pp. 87–94.
[16]M. Castaings, C. Bacon, B. Hosten, and M. V. Predoi, "Finite element predictions for the dynamic response of thermo-viscoelastic material structures," J. Acoust. Soc. Am. **115**, 1125–1133 (2004).
[17]M. Castaings and C. Bacon, "Finite element modeling of torsional wave modes along pipes with absorbing materials," J. Acoust. Soc. Am. **119**, 3741–3751 (2006).
[18]Y.-H. Pao and V. Varatharajulu, "Huygens' principle, radiation conditions, and integral formulas for the scattering of elastic waves," J. Acoust. Soc. Am. **59**, 1361–1371 (1976).
[19]J. E. Gubernatis, E. Domany, and J. A. Krumhansl, "Formal aspects of the theory of the scattering of ultrasound by flaws in elastic materials," J. Appl. Phys. **48**, 2804–2811 (1977).
[20]M. J. S. Lowe and O. Diligent, "Low-frequency reflection characteristics of the s_0 Lamb wave from a rectangular notch in a plate," J. Acoust. Soc. Am. **111**, 64–74 (2002).
[21]M. J. S. Lowe, P. Cawley, J.-Y. Kao, and O. Diligent, "The low-frequency reflection characteristics of the fundamental anti-symmetric Lamb wave a_0 from a rectangular notch in a plate," J. Acoust. Soc. Am. **112**, 2612–2622 (2002).
[22]A. Demma, P. Cawley, M. J. S. Lowe, and A. G. Roosenbrand, "The reflection of the fundamental torsional mode from cracks and notches in pipes," J. Acoust. Soc. Am. **114**, 611–625 (2003).

Scattering of a Bessel beam by a sphere: II. Helicoidal case and spherical shell example

Philip L. Marston[a)]
Department of Physics and Astronomy, Washington State University, Pullman, Washington 99164-2814

(Received 26 March 2008; revised 22 July 2008; accepted 22 July 2008)

In prior work [P. L. Marston, "Scattering of a Bessel beam by a sphere," J. Acoust. Soc. Am. **121**, 753–758 (2007)] the partial wave series for the scattering by a sphere centered on a zero-order Bessel beam was derived. The present work extends the analysis of the far-field scattering to the case of a Bessel beam having an angular dependence on phase. The beam considered is an example of a helicoidal beam where "helicoidal" refers to a type of beam that possesses an axial null and has an azimuthal phase gradient. This type of beam is sometimes also referred to as an acoustic vortex. The beam considered here has a phase ramp equal to the azimuthal angle. In agreement with symmetry arguments given previously, the backward scattering and forward scattering vanish for all frequencies. Some of the resulting modifications of the scattering are illustrated for a rigid sphere and an evacuated steel shell in water. For some directions and choices for the frequency, the calculated scattering by the shell increases when shifting to a helicoidal beam illumination.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2973230]

PACS number(s): 43.40.Fz, 43.25.Qp, 43.30.Ft, 43.35.Ty [ADP] Pages: 2905–2910

I. INTRODUCTION

An ideal Bessel beam satisfies the Helmholtz equation in cylindrical coordinates and propagates without spreading. Marston recently discussed the scattering[1,2] and radiation force[3,4] properties for an ideal acoustic Bessel beam for which the complex pressure is of the form $p_0 \exp[i(\kappa z - \omega t)]J_0(\mu R)$, where J_0 is a zero-order Bessel function, p_0 is a constant, z is the axial coordinate, x and y are transverse coordinates, $R=\sqrt{(x^2+y^2)}$, and $\kappa^2+\mu^2=k^2=(\omega/c)^2$, where c denotes the phase velocity of the fluid. In the prior work, as in the present case, the sphere is taken to be centered on the beam's axis. The present work extends the analysis of the scattering to the case of an isotropic sphere illuminated by the most elementary case of an ideal acoustic helicoidal Bessel beam. For example, the scattering by a rigid sphere and by a spherical shell is numerically evaluated and discussed.

Modern interest in the properties of scalar Bessel beams began with the work of Durnin *et al.*[5,6] in optics which was extended to acoustics by various authors.[7–9] An ideal acoustic Bessel beam is a beam that propagates without spreading. Even though production of such a beam requires an aperture extending over an infinite plane, beams produced with finite aperture transducers can closely resemble a Bessel beam close to the transducer. Consequently scattering calculations for ideal beams may be relevant to understanding scattering that is actually observable. Furthermore, Bessel beams may be superposed to generate more complicated beams and the resulting scattering follows from superposition.[2]

It is widely realized[10–14] that there is a family of exact solutions to the Helmholtz equation where (when placed in the context of acoustics) the spatial part of the complex velocity potential in cylindrical coordinates has an angular dependence and is given by

$$\psi_{Bm}(z,R,\phi) = \psi_0 i^m \exp(i\kappa z + im\phi)J_m(\mu R), \tag{1}$$

where ψ_0 is an amplitude constant, m is an integer, ϕ is an azimuthal angle, J_m is a Bessel function of order m, $\mu/k = \sin\beta$, β is a beam parameter, and the factor i^m is included for convenience. As illustrated in Fig. 1 of Ref. 3, $\beta = \arcsin(\mu/k)$ has the geometric significance of a conic angle. The ordinary Bessel beam solution corresponds to the special case $m=0$. When $m\neq 0$, ψ_{Bm} vanishes on the axis and the angular dependence makes this beam an example of what is sometimes described as a helical,[11] helicoidal,[15–17] or vortex[18,19] beam. While some authors may prefer to describe the beam in Eq. (1) as a "higher order Bessel beam," to avoid possible confusion with solutions of the form $\exp(i\kappa z)\cos(\phi_0+m\phi)J_m(\mu R)$, the present author prefers the term *helicoidal Bessel beam*. Nye and Berry[20] introduced the term "helicoidal wave" in their classification of dislocations in wave fields. Nontrivial helicoidal Bessel solutions do not exist having $\beta=0$ since in that case $\mu=0$ and the $J_m(\mu R)$ factor vanishes.

The present work concerns the scattering of sound by a sphere placed on the axis of such a beam in an ideal fluid for the simplest case, which is $m=1$. Some properties of the scattering are discussed including the suppression of the scattering along the axis of the beam and the backscattering. While not developed here, an understanding of the scattering properties is important for analyzing the acoustic radiation forces[3,4] and torques[15,16] on spheres and may be helpful in the development of "acoustic tweezers" for trapping small objects.[21–23] General properties of the scattering are discussed in Sec. IV and the example of a spherical shell in water is discussed in Sec. V.

[a)]Electronic mail: marston@wsu.edu

II. PLANE-WAVE DECOMPOSITION OF THE BEAM

Following the approach of Durnin *et al.*,[5,6] the helicoidal beam with $m=1$ in Eq. (1) may be described by a superposition of plane waves given by the integral

$$\psi_{B1}(z,R,\phi)=\int_0^{2\pi}\xi_1(z,\phi',R\cos\phi,R\sin\phi)d\phi', \tag{2}$$

where the z axis is the beam axis, x, y, and z are the Cartesian coordinates of the point (z,R,ϕ), $R=\sqrt{(x^2+y^2)}$, $x=R\cos(\phi)$, $y=R\sin(\phi)$, ϕ' is the azimuthal angle of the plane-wave component of the beam, and

$$\xi_1(z,\phi',x,y)=\psi_0(2\pi)^{-1}e^{i\kappa z}\times\exp[i\phi'+i\mu(x\cos\phi'+y\sin\phi')]. \tag{3}$$

The differential component $\xi_1(z,\phi',R\cos\phi,R\sin\phi)\ d\phi'$ represents a plane-wave component in the interval $d\phi'$ characterized by the conic angle $\beta=\arcsin(\mu/k)$ and the azimuthal angle ϕ'. Each differential component depends on the field coordinates (z,R,ϕ). The equivalence of ψ_{B1} in Eq. (2) with ψ_{B1} in Eq. (1) is shown in the Appendix.

III. SCATTERING BY THE SPHERE

Following the procedure described by Marston[1] in the $m=0$ case, the scattering by the sphere is given by integrating the scattering by each plane-wave component. Neglecting the absorption in the surroundings, the far-field scattering for each plane-wave component of the incident wave by a sphere centered at $z=0$ and $R=0$ may be expressed with a complex dimensionless form function f using the differentials

$$d\psi_s=(a/2r)f(ka,\cos\gamma)\exp(ikr)d\psi_{B1}, \tag{4}$$

where a denotes the radius of the sphere, r denotes the distance from the center of the sphere, and γ denotes the scattering angle of the field point relative to the local incident wave vector (shown in Fig. 1 of Marston[1]), and $d\psi_{B1}=\xi_1(0,\phi',0,0)\ d\phi'$, where $\xi_1(0,\phi',0,0)$ is given by evaluating Eq. (3) at the origin. Equation (3) gives $\xi_1(0,\phi',0,0)=\psi_0(2\pi)^{-1}\exp(i\phi')$. Some of the relevant properties of the scattering for the case of plane-wave incidence have been recently reviewed.[1–4] The function f accounts for the dependence of the plane-wave scattering on the frequency and sphere size through the dependence on ka. This function may be expressed in terms of a partial wave series for plane-wave incidence scattering:

$$f(ka,\cos\gamma)=(-i/ka)\sum_{n=0}^{\infty}(2n+1)(s_n-1)P_n(\cos\gamma), \tag{5}$$

where P_n denotes a Legendre polynomial of the indicated argument. Expressions for the function s_n for various types of spheres have also been reviewed.[1–3,24] In the case of an ideal sphere causing no dissipation of acoustic energy, the complex functions s_n are unimodular: $|s_n|=1$. When $|s_n|=1$, s_n becomes $-D_n^*/D_n$, where D_n is the denominator of the partial wave coefficient and $*$ denotes complex conjugation. The superposition of the far-field scattering associated with each plane-wave component of the incident beam gives

$$\psi_s(r,\theta,\phi)=\psi_0(2\pi)^{-1}(a/2r)e^{ikr}\times\int_0^{2\pi}f(ka,\cos\gamma)\exp(i\phi')d\phi'. \tag{6}$$

From Eq. (5) the dependence on $\cos\gamma$ of the nth partial wave is through $P_n(\cos\gamma)$, where

$$P_n(\cos\gamma)=[4\pi/(2n+1)]\sum_{m=-n}^{n}Y_{nm}(\theta,\phi)Y_{nm}^*(\beta,\phi'). \tag{7}$$

This follows from the addition theorem[25] for spherical harmonics Y_{nm} because γ is the angle between a vector in the observer direction (θ,ϕ) and the plane-wave component in direction (β,ϕ'). For the nth partial wave the integration over ϕ' gives a factor

$$I_{nm}=\int_0^{2\pi}\exp(i\phi')Y_{nm}^*(\beta,\phi')d\phi'. \tag{8}$$

It follows from the standard definition of Y_{nm} [Eq. (3.53) of Jackson[25]] that $I_{nm}=0$ for $n=0$ and $I_{nm}=0$ for $m\neq 1$. Furthermore for $m=1$ with $n>0$, $I_{n1}=\pi^{1/2}G_nP_n^{\ 1}(\cos\beta)$, where $G_n=\sqrt{[(2n+1)(n-1)!/(n+1)!]}$, $P_n^{\ 1}$ is an associated Legendre function, and the factorial function is such that $(n-1)!=1$ for $n=1$. Since only the $m=1$ term remains from Eq. (7), application of the definition of $Y_{nm}(\theta,\phi)$ gives

$$\psi_s(r,\theta,\phi)=\psi_0(a/2r)e^{ikr}F(ka,\theta,\phi), \tag{9}$$

where the dimensionless complex form function is

$$F=(-i/ka)\sum_{n=1}^{\infty}(s_n-1)E_n(\theta,\phi), \tag{10}$$

where $E_n=(G_n)^2\ P_n^{\ 1}(\cos\theta)\ P_n^{\ 1}(\cos\beta)\ \exp(i\phi)$. It follows that

$$E_n=[(2n+1)/n(n+1)]P_n^{\ 1}(\cos\theta)P_n^{\ 1}(\cos\beta)e^{i\phi}. \tag{11}$$

While not central to the current discussion, the approach here may be generalized to give the scattering in the near field. In the $m=0$ case, calculations based on the finite element method support the validity of this approach.[26] In the computations given here, the series in Eq. (10) was terminated for n far in excess of $1.6ka$, a condition for convergence.[4]

IV. PROPERTIES OF THE SCATTERING

The main result of this analysis is Eqs. (9)–(11) which express the far-field scattering by an isotropic sphere centered on a Bessel beam, with $m=1$ from Eq. (1). Since at a fixed frequency the complex pressure is proportional to the velocity potential, ψ_0 in Eqs. (1) and (9) may be jointly replaced by a pressure constant p_0. In a way that is analogous to the prior results[1–4] for illumination by a Bessel beam with $m=0$, each partial wave is proportional to (s_n-1), where (s_n-1) is the same function used in the plane-wave illumination case. The denominator D_n of s_n contains the information concerning the complex frequency "pole" for each resonance.[24,27] The poles are complex ka roots of $D_n(ka)=0$. Each pole location (often referred to simply as a reso-

nance) depends on the partial wave index n and is a property of the sphere and of the fluid loading but not of the type of acoustic beam or location of the sound source used to excite the scattering.[2] An important feature of the scattering by a sphere centered on the helicoidal Bessel beam is that the coupling to the monopole ($n=0$) partial wave is *completely suppressed* for all frequencies. This is evident from Eq. (10) in the $m=1$ case and it is a property generalized to beams with $m>1$ from symmetry. The suppressed coupling is expected to be especially important in the scattering by gas bubbles because ordinarily the $n=0$ partial wave contribution dominates the low frequency behavior.[28] This suppressed coupling will no longer be present if the sphere is offset from the center of the beam.

With the sphere centered on the beam, the dependence of the scattering on the beam parameter β is through the factor $P_n^1(\cos\beta)$ in Eq. (11) in the $m=1$ case and in the $m=0$ case through a factor[1–4] $P_n(\cos\beta)$. It follows that, as for the previous case of a beam with[2] $m=0$, it is possible to select the beam parameter β so as to suppress the coupling to the nth partial wave. In the $m=1$ case this is achieved by taking β to be a nontrivial root of $P_n^1(\cos\beta)=0$. From chiral symmetry this suppression will also be true for beams given by Eq. (1) having $m=-1$. That property is also expected because[25] $P_n^{-m}(\cos\beta)$ is proportional to $P_n^m(\cos\beta)$.

Consider now the angular scattering properties implied by Eqs. (9)–(11). The on-axis scattering is given by taking $\theta=0$ or π. That scattering vanishes since for $\theta=0$ or π, $P_n^1(\cos\theta)$ vanishes. The on-axis scattering vanishes for all values of the dimensionless frequency ka. Hefner and Marston[16] previously noted (from symmetry) that the "backscattering amplitude vanishes" for axisymmetric targets on the axis of a general helicoidal wave. The vanishing of the forward scattering for axisymmetric targets on the axis of a general helicoidal wave also follows from symmetry. In that case the illumination is such that the forward diffraction peak, which is usually large[1] when ka is large, may be suppressed. In the usual case of plane-wave illumination, Fig. 2 of Marston[1] illustrates the ordinary forward diffraction peak for rigid and soft spheres having a $ka=15$.

To illustrate the suppression of the forward and the backward scattering, Eq. (10) was evaluated for a rigid sphere. In the case of a rigid sphere the function $s_n(ka)$ is[1,24] $-h_n^{(2)}(ka)'/h_n^{(1)}(ka)'$, where h_n is a spherical Hankel function of the indicated kind and primes denote differentiation with respect to the indicated argument. Figure 1 shows $|F|$ as a function of the scattering angle θ for a rigid sphere with $ka=15$. Three values of β are shown: 15° (short dashes), 30° (long dashes), and 45° (solid curve). For each curve the scattering is suppressed at 0° and at 180°. Each curve has a peak in $|F|$ centered near $\theta=\beta$. This peak may be interpreted as associated with the superposition of the local forward scattering of the individual plane-wave components of the incident helicoidal beam. Each curve displays structure in the backward hemisphere that becomes narrower for *increasing* β.

Though not emphasized here, helicoidal electromagnetic[29,30] and acoustic[15,16] waves have been predicted to transport orbital angular momentum. Electromagnetic as well as acoustic helicoidal waves both have an azimuthal dependence proportional to $\exp(im\phi)$ and it is the orbital angular momentum associated with that dependence that is relevant to the acoustical angular momentum transport.[16] For electromagnetic waves carrying angular momentum, symmetric illumination of an isotropic sphere causes a radiation torque when there is dissipation within the sphere associated with the absorption of electromagnetic energy.[31–33] It has been anticipated that there will be an analogous acoustical torque associated with the dissipation of mechanical energy in the scattering by an isotropic sphere.[16] Lekner,[34] however, suggested that his analysis of angular momentum flux for a superposition of Bessel helicoidal beams having a fixed $m\neq 0$ may differ from that of Hefner and Marston.[15,16] Lekner's concern is resolved when it is remembered that ideal Bessel beams described by Eq. (1) *do not spread* as they propagate. Recent experiments confirm the presence of radiation torques on circular disks illuminated by propagating acoustical helicoidal waves.[35,36]

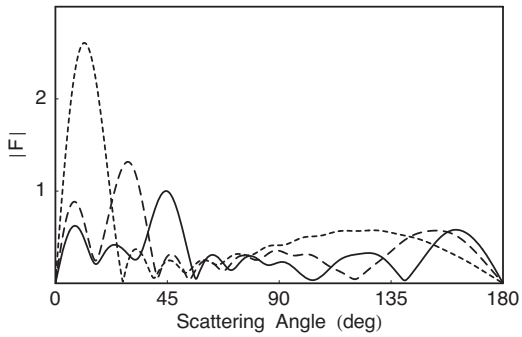


FIG. 1. The form function modulus $|F|$ for scattering by a rigid sphere illuminated by a helicoidal Bessel beam having the angular index $m=1$. The abscissa is the scattering angle θ, where θ is measured relative to the axis of the beam. The dimensionless frequency is $ka=15$, where a is the radius of the sphere. The curves have different beam parameters: $\beta=15°$ (short dashes), 30° (long dashes), and 45° (solid curve). The curves show that relative to the beam axis, the scattering is suppressed in the forward and backward directions.

Marston[2] previously noted how to compute the scattering by a sphere centered on a superposition of Bessel beams having $m=0$. That method, when applied to Eqs. (1) and (9)–(11), may be generalized to give the scattering by a sphere centered on a superposition of helicoidal Bessel beams having $m=1$.

V. SCATTERING BY AN EVACUATED STEEL SHELL

Marston[2] discussed the effects of ordinary Bessel beam illumination on the scattering by an evacuated stainless steel shell in water. The generalization to the case of helicoidal Bessel beam is considered here. As in the prior example[2] the shell properties are selected as follows: the ratio of the inner radius to the outer radius is taken to be 0.95 and the material and water properties were as follows: longitudinal and transverse wave speeds $c_L=5790$ m/s, $c_T=3100$ m/s, and density $\rho_E=7.90$ g/cm^3 for steel and $c=1482$ m/s and density $\rho=1.00$ g/cm^3 for water. These properties were selected to be similar to those of a shell used in underwater acoustic measurements.[37] Since the dissipation of mechanical energy by the shell is expected to be small when ka is small, dissi-

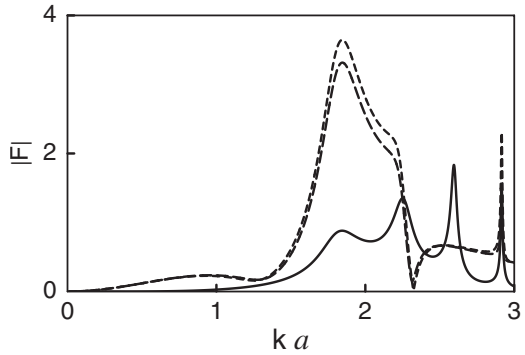


FIG. 2. The calculated modulus of the dimensionless form function F for the scattering by an evacuated steel shell in water plotted as a function of the dimensionless frequency ka. The scattering angle θ is fixed at 149.444…° [which is a root of $P_4(\cos\theta)=0$]. The curve with short dashes is for plane-wave illumination. The curve with long dashes and the solid curve are for Bessel beam illumination having $m=0$ and $m=1$, respectively. In both Bessel beam cases the beam parameter β is 15°. Note that near $ka=2.595$, the scattering is greatly increased in the case of $m=1$ in comparison with the plane wave and $m=0$ cases.

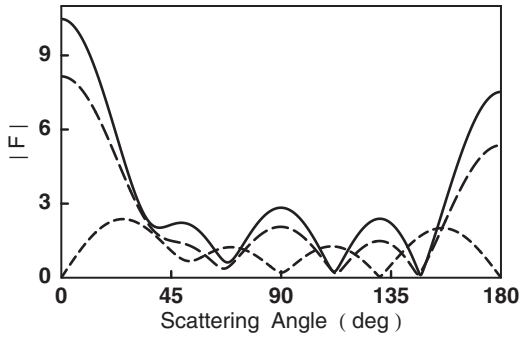


FIG. 3. Calculated angular scattering patterns for the shell considered in Fig. 2 for $ka=2.595$ which is the frequency of a resonance having a mode index $n=4$. The incident waves are a plane wave (solid curve) and Bessel beams with $\beta=15°$ and $m=0$ (long dashes) and $\beta=15°$ and $m=1$ (short dashes). The forward and backward scattering is suppressed in the helicoidal case. In some angular regions the scattering increases by changing from plane wave or ordinary ($m=0$) Bessel beam illumination to the case of a helicoidal beam having $m=1$.

pation is completely neglected here so that $s_n=-D_n^*/D_n$, where the function $D_n(ka)$ is reviewed by Marston[2] for the present case of an evacuated shell. Resonance features of the low-lying partial waves are associated with subsonic guided waves and have some similarities with low frequency modes of some other elastic spherical targets.[2,4,38] Figure 2 shows the dependence of the dimensionless form function magnitude $|F|$ on ka for this shell for three situations where the scattering angle θ is given by the root of $P_4(\cos\theta)=0$ closely approximated by 149.444°. This scattering angle was chosen to illustrate a situation where illumination by a helicoidal Bessel beam can cause an increase in the scattering relative to some alternative forms of illumination. The curve having short dashes is the case of plane-wave illumination where the vertical axis is $|f(ka,\cos\theta)|$ from Eq. (5). The curve having long dashes is for Bessel beam illumination with $m=0$ and $\beta=15°$. In this case $|F|$ is given by Eq. (9) of Marston.[1] The solid curve is for a helicoidal Bessel beam having $m=1$ and $\beta=15°$ and is given by Eq. (10). It is helpful to compare the solid curve with the solid curve of Fig. 3 of Marston[2] which shows $|f(ka,\cos 180°)|$, the backscattering amplitude with plane-wave illumination. The peak locations are similar for both of those solid curves and (as previously noted[2]) the peaks are associated with partial waves having $n=2$, 3, 4, and 5 (numbering from left to right). The peak for $n=4$ is at $ka=2.595$. This peak is completely suppressed for the dashed curves in Fig. 2 because for both of those curves the $n=4$ partial wave has a factor $P_4(\cos\theta)$ which vanishes for the selected value of θ. The scattering increases in the helicoidal beam case since the dependence on θ for the $n=4$ partial wave is replaced by $P_n^1(\cos\theta)=2.3978$ at $\theta=149.444°$.

Figure 3 shows angular scattering patterns for the same shell evaluated at $ka=2.595$ so as to strongly excite the $n=4$ partial wave. The solid curve is for plane-wave illumination and is given by $|f(ka,\cos\theta)|$ from Eq. (5). The scattering is not exactly proportional to $P_4(\cos\theta)$ since there are noticeable contributions from other partial waves, especially in the forward hemisphere. The scattering near 149.444° [where $P_4(\cos\theta)$ vanishes] is quite weak. That is also the case for the curve having long dashes, which is for a Bessel beam with $m=0$ having $\beta=15°$ and is given by Eq. (9) of Marston.[1] The curve having short dashes is the helicoidal beam case, Eq. (10), with $\beta=15°$. The angular dependence is approximately that of the function $|P_4^1(\cos\theta)|$, though deviations from that behavior are evident since the pattern is not perfectly symmetric around $\theta=90°$. For some narrow ranges of scattering angles the helicoidal scattering amplitude exceeds the amplitudes in the other two cases.

For additional insight into the predicted increase in scattering in certain angular regions with helicoidal Bessel beam illumination, consider the dependence of the $J_m(\mu R)$ factor in Eq. (1) for the two Bessel beam cases considered in Fig. 3. For these cases $\mu R=(R/a)\ ka\sin\beta$, with $ka=2.595$ and $\beta=15°$. The factors $J_m(\mu R)$ are plotted in Fig. 4 for $m=0$ (dashed curve) and $m=1$ (solid curve). For a fixed beam amplitude parameter ψ_0, note that in the region of the shell ($R\leq a$) the local beam amplitude is far less in the $m=1$ case than for $m=0$. In all three cases shown in Fig. 3 the incident wave couples into shell modes with $n=4$. Even with rela-

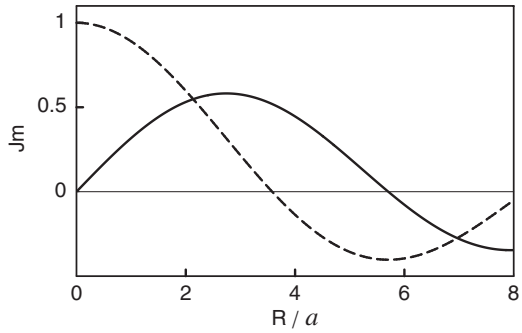


FIG. 4. The Bessel function factor $J_m(\mu R)$ in the local beam amplitude, Eq. (1), plotted as a function of R/a, where R is the transverse distance from the beam axis and a is the sphere radius. The plots are for $ka=2.595$ and $\beta=15°$ for beams with $m=0$ (dashed curve) and $m=1$ (solid curve). These are the conditions of the beams considered in Fig. 3. This plot shows that for the same value of ψ_0 in Eq. (1), the incident beam amplitude in the region of the sphere is significantly reduced in the $m=1$ case (the solid curve) in comparison to the $m=0$ case (the dashed curve). Nevertheless, Fig. 3 shows that for this ka the scattering into some angular regions is increased in the $m=1$ case for the example of the shell considered.

tively weak local illumination evident in Fig. 4, the reason why the scattering is greater in certain angular regions for the $m=1$ case is that the $\exp(i\phi)$ dependence of that beam gives rise to a superposition of $m=1$ modes having the form of an angular traveling wave (as opposed to a standing wave) on the shell. The radiation pattern in the traveling wave case, which is proportional to $P_4^1(\cos\theta)$, is quite different from that of the ordinary pattern proportional to $P_4(\cos\theta)$.

The $\exp(im\phi)$ dependence of helicoidal illumination is an example of what Ho and Felsen[39] termed illumination with an azimuthal "phase gradient." While their analysis of the response of shells illuminated by sources having azimuthal phase gradients is not directly applicable to the situation considered here, the significant modifications of the scattering by a shell illustrated here for $m=1$ are not surprising given the range of modifications of shell responses considered by Ho and Felsen.[39]

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research. Dr. David B. Thiessen and Likun Zhang (both at Washington State University) and an anonymous reviewer suggested helpful improvements to the manuscript.

APPENDIX: DERIVATION OF PLANE-WAVE DECOMPOSITION

Schelkunoff, in his textbook[40] (based on his lectures from the 1930s), expressed the geometric relationship between waves in a circular cylindrical waveguide and plane waves by using a cone of wavevectors. The approach given below was motivated in part by his analysis. The wavevector for a single component of the angular spectrum may be designated as $\mathbf{k}(\beta,\phi')$ which has components $(k_x,k_y,k_z)=(\mu\cos\phi',\mu\sin\phi',\kappa)$, where $\kappa=k\cos\beta$ and $\mu=k\sin\beta$. Consider the following angular superposition

$$I_m(x,y,z)=\int_0^{2\pi}\exp[im\phi'+i\mathbf{k}(\beta,\phi')\cdot\mathbf{r}]d\phi', \tag{A1}$$

where the components of $\mathbf{r}$ are $(x,y,z)=(R\cos\phi,R\times\sin\phi,z)$. It follows that

$$I_m=e^{i\kappa z}\int_0^{2\pi}\exp[im\phi'+i\mu R\cos(\phi-\phi')]d\phi'.$$

Let $\phi''=\phi'-\phi$ so that $I_m=\exp(i\kappa z+im\phi)\ L_m$, where

$$L_m=\int_{-\phi}^{2\pi-\phi}\exp(im\phi''+i\mu R\cos\phi'')d\phi''.$$

This may be evaluated using the following standard expansion theorem:[41]

$$\exp(i\mu R\cos\phi'')=\sum_{m'=-\infty}^{\infty}i^{m'}\exp(im'\phi'')J_{m'}(\mu R), \tag{A2}$$

which gives $L_m=i^{-m}2\pi J_{-m}(\mu R)=i^m 2\pi J_m(\mu R)$, so that in Eq. (1),

$$\psi_{Bm}=\psi_0(2\pi)^{-1}I_m(x,y,z). \tag{A3}$$

Equations (2) and (3) follow from Eqs. (A1) and (A3) in the case $m=1$.

[1]P. L. Marston, "Scattering of a Bessel beam by a sphere," J. Acoust. Soc. Am. **121**, 753–758 (2007).
[2]P. L. Marston, "Acoustic beam scattering and excitation of sphere resonance: Bessel beam example," J. Acoust. Soc. Am. **122**, 247–252 (2007).
[3]P. L. Marston, "Axial radiation force of a Bessel beam on a sphere and direction reversal of the force," J. Acoust. Soc. Am. **120**, 3518–3524 (2006).
[4]P. L. Marston, "Negative axial radiation forces on solid spheres and shells in a Bessel beam," J. Acoust. Soc. Am. **122**, 3162–3165 (2007).
[5]J. Durnin, "Exact solutions for nondiffracting beams. I. The scalar theory," J. Opt. Soc. Am. A **4**, 651–654 (1987).
[6]J. Durnin, J. J. Miceli, Jr., and J. H. Eberly, "Diffraction-free beams," Phys. Rev. Lett. **58**, 1499–1501 (1987).
[7]J. A. Campbell and S. Soloway, "Generation of a nondiffracting beam with frequency-independent beamwidth," J. Acoust. Soc. Am. **88**, 2467–2477 (1990).
[8]J.-Y. Lu and J. F. Greenleaf, "Ultrasonic nondiffracting transducer for medical ultrasonics," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 438–447 (1990).
[9]P. L. Marston, "Geometrical and catastrophe optics methods in scattering," Phys. Acoust. **21**, 1–234 (1992), see Sec. 2.15, pp. 44–47.
[10]G. Indebetouw, "Nondiffracting optical fields: Some remarks on their analysis and synthesis," J. Opt. Soc. Am. A **6**, 150–152 (1989).
[11]P. L. Overfelt, "Scalar optical beams with helical symmetry," Phys. Rev. A **46**, 3516–3522 (1992).
[12]C. J. Chapman, "Caustics in cylindrical ducts," Proc. R. Soc. London, Ser. A **455**, 2529–2548 (1999).
[13]C. Paterson and R. Smith, "Higher-order Bessel waves produced by axicon-type computer-generated holograms," Opt. Commun. **124**, 121–130 (1996).
[14]J. Arlt and K. Dholakia, "Generation of high-order Bessel beams by use of an axicon," Opt. Commun. **177**, 297–301 (2000).
[15]B. T. Hefner and P. L. Marston, "Acoustical helicoidal waves and Laguerre-Gaussian beams: Applications to scattering and to angular momentum transport," J. Acoust. Soc. Am. **103**, 2971 (1998); *Proceedings of the Joint Meeting of the 16th International Congress on Acoustics and the 137th Meeting of the Acoustical Society of America* (Acoustical Society of America, New York, 1998), pp. 1921–1922.
[16]B. T. Hefner and P. L. Marston, "An acoustical helicoidal wave transducer with applications for the alignment of ultrasonic and underwater systems," J. Acoust. Soc. Am. **106**, 3313–3316 (1999).
[17]S. Gspan, A. Meyer, S. Bernet, and M. Ritsch-Marte, "Optoacoustic generation of a helicoidal ultrasonic beam," J. Acoust. Soc. Am. **115**, 1142–1146 (2004).
[18]J.-L. Thomas and R. Marchiano, "Pseudo angular momentum and topological charge conservation for nonlinear acoustical vortices," Phys. Rev. Lett. **91**, 244302 (2003).
[19]R. Marchiano and J.-L. Thomas, "Synthesis and analysis of linear and nonlinear acoustic vortices," Phys. Rev. E **71**, 066616 (2005).
[20]J. F. Nye and M. V. Berry, "Dislocations in wave trains," Proc. R. Soc. London, Ser. A **336**, 165–190 (1974).
[21]J. R. Wu, "Acoustical tweezers," J. Acoust. Soc. Am. **89**, 2140–2143 (1991).
[22]J. W. Lee and K. K. Shung, "Radiation forces exerted on arbitrarily located sphere by acoustic tweezer," J. Acoust. Soc. Am. **120**, 1084–1094 (2006).
[23]F. G. Mitri, "Acoustic radiation force on a sphere in standing and quasi-standing zero-order Bessel beam tweezers," Ann. Phys. **323**, 1604–1620 (2008).
[24]L. Flax, G. C. Gaunaurd, and H. Überall, "Theory of resonance scattering," Phys. Acoust. **10**, 191–294 (1981).
[25]J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999), Sec. 3.5, pp. 107–111.
[26]D. B. Thiessen and P. L. Marston, "Negative radiation forces on spheres illuminated by Bessel beams: Modeling using finite elements," J. Acoust. Soc. Am. **122**, 3025(A) (2007).
[27]R. H. Hackman, "Acoustic scattering from elastic solids," Phys. Acoust. **22**, 1–194 (1993).
[28]A. Prosperetti, "Thermal effects and damping mechanisms in the forced

radial oscillations of gas bubbles in liquids," J. Acoust. Soc. Am. **61**, 17–27 (1977).
[29]L. Allen, M. J. Padgett, and M. Babiker, "The orbital angular momentum of light," Prog. Opt. **39**, 291–372 (1999).
[30]L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman, "Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes," Phys. Rev. A **45**, 8185–8189 (1992).
[31]P. L. Marston and J. H. Crichton, "Radiation torque on a sphere caused by a circularly polarized electromagnetic wave," Phys. Rev. A **30**, 2508–2516 (1984).
[32]J. H. Crichton and P. L. Marston, "The measurable distinction between the spin and orbital angular momenta of electromagnetic radiation," Electron. J. Differ. Equations **04**, 37–50 (2000); open access journal online at http://ejde.math.unt.edu (Last viewed February, 2008).
[33]T. A. Nieminen, V. L. Y. Loke, A. B. Stilgoe, G. Knöner, A. M. Branczyk, N. R. Heckenberg, and H. Rubinsztein-Dunlop, "Optical tweezers computational toolbox," J. Opt. A, Pure Appl. Opt. **9**, S196–S203 (2007).
[34]J. Lekner, "Acoustic beams with angular momentum," J. Acoust. Soc. Am. **120**, 3475–3478 (2006).
[35]K. Volke-Sepúlveda, A. O. Santillán, and R. R. Boullosa, "Transfer of angular momentum to matter from acoustical vortices in free space," Phys. Rev. Lett. **100**, 024302 (2008).
[36]K. D. Skeldon, C. Wilson, M. Edgar, and M. J. Padgett, "An acoustic spanner and its associated rotational Doppler shift," New J. Phys. **10**, 013018 (2008).
[37]H. J. Simpson, B. H. Houston, and R. Lin, "Laboratory measurements of sound scattering from a buried sphere above and below the critical angle," J. Acoust. Soc. Am. **113**, 39–42 (2003).
[38]K. J. Diercks and R. Hickling, "Echoes from hollow aluminum spheres in water," J. Acoust. Soc. Am. **41**, 380–393 (1967).
[39]J. M. Ho and L. B. Felsen, "Fully three-dimensional exact and ray asymptotic formulation of the characteristic wave fields on a spherical shell surface," J. Acoust. Soc. Am. **95**, 265–285 (1994).
[40]S. A. Schelkunoff, *Electromagnetic Waves* (Van Nostrand, New York, 1943), pp. 410–412.
[41]A. L. Fetter and J. D. Walecka, *Theoretical Mechanics of Particles and Continua* (Dover, Mineola, NY, 2003), pp. 336–338.

Generating nonstationary multisensor signals under a spatial coherence constraint

Emanuël A. P. Habets[a)]
School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel, and Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel

Israel Cohen[b)]
Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel

Sharon Gannot[c)]
School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel

(Received 27 June 2008; revised 20 August 2008; accepted 20 August 2008)

Noise fields encountered in real-life scenarios can often be approximated as spherical or cylindrical noise fields. The characteristics of the noise field can be described by a spatial coherence function. For simulation purposes, researchers in the signal processing community often require sensor signals that exhibit a specific spatial coherence function. In addition, they often require a specific type of noise such as temporally correlated noise, babble speech that comprises a mixture of mutually independent speech fragments, or factory noise. Existing algorithms are unable to generate sensor signals such as babble speech and factory noise observed in an arbitrary noise field. In this paper an efficient algorithm is developed that generates multisensor signals under a predefined spatial coherence constraint. The benefit of the developed algorithm is twofold. Firstly, there are no restrictions on the spatial coherence function. Secondly, to generate M sensor signals the algorithm requires only M mutually independent noise signals. The performance evaluation shows that the developed algorithm is able to generate a more accurate spatial coherence between the generated sensor signals compared to the so-called image method that is frequently used in the signal processing community. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2987429]

PACS number(s): 43.50.Ed, 43.60.Cg [EJS] Pages: 2911–2917

I. INTRODUCTION

A spherical noise field has been shown to be a reasonable model for a number of practical noise fields that can be found in, for example, an office or a car.[1] Cylindrical noise fields are especially useful when, for example, the ceiling and floor in an enclosure are covered with a highly absorbing material.[1] Spherical and cylindrical noise fields are also known as three-dimensional (3D) and two-dimensional (2D) diffuse noise fields, respectively. Researchers in the signal processing community often require sensor signals that result from these noise fields for simulation purposes, e.g., for (superdirective) beamforming,[2–5] adaptive noise cancellation,[6,7] and source localization. In some cases it might even be desired to generate sensor signals that exhibit a specific spatial coherence function, e.g., based on a specific measurement or condition.

It is often assumed that the noise field is (i) spatially homogeneous, i.e., the physical properties of the sound do not depend on the absolute position of the sensor, (ii) isotropic, i.e., the physical properties of the sound are the same in any direction of measurement, and (iii) time invariant.[1,8] The sensor signals acquired in 2D and 3D diffuse noise fields can be generated using a number of independent noise sources that are uniformly spaced on a cylinder or sphere, respectively.[6,8,9] Recently, we have developed an efficient algorithm to generate sensor signals acquired in noise fields that satisfy the above assumptions.[10]

In many cases noise comprises a mixture of independent speech fragments, also known as babble speech, or factory noise. Babble speech can be used to model the background noise encountered in a multitalker environment such as a restaurant or cafeteria. In such a case the short-term power spectral densities (PSDs) of the sensor signals vary in space and time. This relaxes some of the prior assumptions. The algorithm in Ref. 10 can be used to generate stationary and nonstationary signals. However, the algorithm requires a large number of independent noise signals. Furthermore, the spatial coherence depends on the positions of the noise sources. Therefore, it is difficult to obtain an arbitrary spatial coherence. In this paper, we develop an efficient algorithm that generates nonstationary sensor signals under a predefined spatial coherence constraint. The benefit of the developed algorithm is twofold. Firstly, there are no restrictions on the spatial coherence function, allowing the use of an arbitrary or measured spatial coherence function. For example, generated sensor signals with slightly different spatial coherence functions can be used to analyze the robustness and the sensitivity of acoustic signal processing algorithms. Secondly, to generate M sensor signals the developed algo-

[a)]Electronic mail: e.habets@ieee.org. URL: http://home.tiscali.nl/ehabets
[b)]Electronic mail: icohen@ee.technion.ac.il. URL: http://www.ee.technion.ac.il/Sites/People/IsraelCohen
[c)]Electronic mail: gannot@eng.biu.ac.il. URL: http://www.eng.biu.ac.il/~gannot

rithm only requires M mutually independent noise signals, which is a number that is significantly smaller than the number of noise signals that is used by other algorithms.[9,10] The developed algorithm generates the sensor signals in two steps. Firstly, we generate a set of mutually independent noise signals. The cardinality of this set is equal to the number of sensors. Secondly, the noise signals are filtered and mixed such that the obtained sensor signals exhibit a predefined spatial coherence. The filtering and mixing can be performed in the frequency domain or in the time-frequency domain such that these operations reduce to an instantaneous mixing.

The remainder of this paper is organized as follows: In Sec. II, we formulate the problem of generating sensor signals under a predefined spatial coherence. The sensor signals are expressed as instantaneous mixing of mutually independent noise signals. In Sec. III, we generate the mutually independent noise signals and in Sec. IV we compute the instantaneous mixing matrix. In Sec. V we summarize the algorithm and discuss its computational complexity. In Sec. VI, we evaluate the performance in different noise fields by comparing the spatial coherence of the generated sensor signals with the theoretical spatial coherence.

II. PROBLEM FORMULATION

Our objective is to generate M sensor signals with a predefined spatial coherence. The M sensor positions, relative to the first sensor position, are stacked into a matrix $\mathbf{P}$, such that

$$\mathbf{P} = \begin{bmatrix} 0 & x_2 & \cdots & x_M \\ 0 & y_2 & \cdots & y_M \\ 0 & z_2 & \cdots & z_M \end{bmatrix}. \tag{1}$$

The Euclidian distance d_{pq} between the pth and the qth sensor is given by

$$d_{pq} = \|\mathbf{P}_p - \mathbf{P}_q\|_2, \tag{2}$$

where $\mathbf{P}_p$ denotes the pth column of the matrix $\mathbf{P}$.

Let us denote the PSD of the pth sensor signal by $\Phi_{pp}(\omega)$, where ω denotes the angular frequency. The cross-PSD between the pth and the qth sensor is denoted by $\Phi_{pq}(\omega)$.

The assumption that the noise field is homogeneous can be formulated as

$$\Phi_{pp}(\omega) = \Phi(\omega), \quad \forall\, p \in \{1, \ldots, M\}. \tag{3}$$

The spatial coherence between the pth and the qth sensor is defined as[8]

$$\gamma_{pq}(\omega) = \frac{\Phi_{pq}(\omega)}{\sqrt{\Phi_{pp}(\omega)\Phi_{qq}(\omega)}}. \tag{4}$$

In a spherically isotropic noise field the spatial coherence function is given by[8,11]

$$\gamma_{pq}(\omega) = \frac{\sin(\omega d_{pq}/c)}{\omega d_{pq}/c}, \tag{5}$$

where d_{pq} denotes the distance between the pth and qth sensors, and c denotes the sound velocity in ms^{-1}. Another well-known noise field is a cylindrically isotropic noise field. The spatial coherence function is then given by[12]

$$\gamma_{pq}(\omega) = J_0(\omega d_{pq}/c), \tag{6}$$

where $J_0(\cdot)$ is a zero-order Bessel function of the first kind.

We propose to filter and mix M mutually independent noise signals to generate M sensor signals that exhibit a predefined spatial constraint. Since the spatial coherence is defined in the frequency domain it is preferred to work in this domain. The filtering and mixing can efficiently be performed in the frequency domain and in the time-frequency domain. Here, we work in the short-time Fourier transform (STFT) domain. The frame index is denoted by ℓ and the discrete angular frequencies are denoted by ω_k, where $(k \in \{0, \ldots, K/2-1\})$ and K is the frame length of the STFT. It should be noted that the spatial coherence can either be (slowly) time varying or time invariant. While the developed algorithm can be used to simulate a (slowly) time varying spatial coherence, we assume that the predefined spatial coherence is time invariant.

Let us define a matrix $\tilde{\mathbf{\Gamma}}(\omega_k)$ for each ω_k that consists of the predefined spatial coherence values

$$\tilde{\mathbf{\Gamma}}(\omega_k) = \begin{bmatrix} \tilde{\gamma}_{11}(\omega_k) & \tilde{\gamma}_{12}(\omega_k) & \cdots & \tilde{\gamma}_{1M}(\omega_k) \\ \tilde{\gamma}_{21}(\omega_k) & \tilde{\gamma}_{22}(\omega_k) & \cdots & \tilde{\gamma}_{2M}(\omega_k) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\gamma}_{M1}(\omega_k) & \tilde{\gamma}_{M2}(\omega_k) & \cdots & \tilde{\gamma}_{MM}(\omega_k) \end{bmatrix}. \tag{7}$$

Examples of $\tilde{\gamma}_{pq}(\omega k)$ are given by Eqs. (5) and (6).

The prerequisites for the sensor signals can be summarized as follows.

(1) The spatial coherence $\gamma_{pq}(\omega k)$ between the pth and the qth sensor should be equal to a predefined spatial coherence $\tilde{\gamma}_{pq}(\omega k)$.
(2) The PSDs of the sensor signals should be equal.

Let us define a vector that consists of the STFT coefficients of the sensor signals $\mathbf{X}(\ell, \omega_k) = [X_1(\ell, \omega_k), \ldots X_M(\ell, \omega_k)]^T$ and the noise signals $\mathbf{N}(\ell, \omega_k) = [N_1(\ell, \omega_k), \ldots, N_M(\ell, \omega_k)]^T$. The STFT coefficients of the sensor signals are obtained by instantaneous mixing of the STFT coefficients of the noise signals, i.e.,

$$\mathbf{X}(\ell, \omega_k) = \mathbf{C}^H(\omega_k)\mathbf{N}(\ell, \omega_k), \tag{8}$$

where $\mathbf{C}(\omega_k)$ denotes the mixing matrix and $(\cdot)^H$ is the Hermitian operation.

In Sec. III, we generate M mutually independent noise signals $N_p(\ell, \omega_k)$ $p \in \{1, 2, \ldots, M\}$. In Sec. IV, we show how to compute the mixing matrix $\mathbf{C}(\omega_k)$ such that the spatial coherence of the sensor signals equals Eq. (7).

III. GENERATE NOISE SIGNALS

In this section we provide two alternatives to generate the M mutually independent noise signals. In the following we assume that the column vectors of the mixing matrix have equal norm. If the short-term PSDs of the noise signals are equal then the short-term PSDs of the sensor signals are equal. Hence, the PSDs of the sensor signals are consistent with the PSDs observed in a homogeneous noise field.

A. Perfectly homogeneous

We generate M mutually independent noise signals with short-term PSD $\Phi(\ell,\omega_k)$. The short-term PSD of the noise signal can be time varying or time invariant, and spectrally white or colored.

We generate the STFT coefficients of the pth noise signal as

$$N_p(\ell,\omega_k) = \sqrt{\Phi(\ell,\omega_k)}\exp(i\pi D_p(\ell,\omega_k)), \tag{9}$$

where $D_p(\ell,\omega_k)$ for $\ell \in \{1,2,\ldots\}$ denotes a random signal with uniform distribution $(D_p(\ell,\omega_k) \sim U[-1,1])$. Now M mutually independent noise signals can be generated using mutually independent signals $D_p(\ell,\omega_k)$ for $p \in \{1,2,\ldots,M\}$, and the same $\Phi(\ell,\omega_k)$.

The short-term PSD $\Phi(\ell,\omega_k)$ can represent the long-term PSD of speech or factory noise, such that $\Phi(\ell,\omega_k) = \Phi(\omega_k)$. Alternatively, the short-term PSD $\Phi(\ell,\omega_k)$ can be equal to the short-term PSD of a given babble speech or factory noise signal. When babble speech or factory noise is used it should be noted that the original phase spectrum is destructed. Therefore, the resulting babble speech or factory noise signals do not sound like the original signals anymore. In the case when we use babble speech, the resulting noise signals are also known as *babble noise*.

B. Approximately homogeneous

We generate M mutually independent signals that consist of babble speech or factory noise. In the following we assume that these noise signals are continuous, i.e., there are no periods of silence. As mentioned before it is important that the noise signals have the same power. Therefore, we normalize the power of the noise signals $2 \leqslant p \leqslant M$ such that they have the same power as the first noise signal. The noise signals are first transformed into the STFT domain to construct the STFT coefficients $N_p(\ell,\omega_k)$ for $p \in \{1,2,\ldots,M\}$. When these signals are used to generate the sensor signals, the simulated noise field is not completely homogeneous since the short-term PSDs of $N_p(\ell,\omega_k)$ and $N_q(\ell,\omega_k)$ are not equal for $p \neq q$. However, in some cases, the long-term PSDs are approximately equal. For example, due to the nature of speech this is the case for babble speech signals. Therefore, only small fluctuations in the long-term PSDs are expected. Hence, the simulated noise field is approximately homogeneous.

It is important to note that when these noise signals are mixed to generate the desired spatial coherence the resulting sensor signals sound like a mixture of M babble speech or factory noise signals; this is not the case when we use the noise signals suggested in Sec. III A.

IV. GENERATING COHERENCE

In this section we determine the mixing matrix $\mathbf{C}(\omega_k)$ that generates the desired spatial coherence between the sensor signals.

In terms of the mixing matrix $\mathbf{C}(\omega_k)$ we can formulate the prerequisites as follows.

(1) The inner product between the pth and the qth column vector should be equal to $\tilde{\gamma}_{pq}(\omega_k)$.
(2) The norm of the column vectors should be equal to 1.

A. Two sensors

In this section we develop an efficient technique to generate two sensor signals with a predefined spatial coherence. Here we assume that the predefined spatial coherence between the first and the second sensor $\tilde{\gamma}_{12}(\omega_k)$ is real valued.

Let us define the following mixing matrix:

$$\mathbf{C}(\omega_k) = \begin{bmatrix} 1 & \sin(\alpha_{12}(\omega_k)) \\ 0 & \cos(\alpha_{12}(\omega_k)) \end{bmatrix}. \tag{10}$$

It can easily be verified that the norm of the column vectors equals to 1, and that the inner product equals to $\sin(\alpha_{12}(\omega_k))$. Hence, the spatial coherence $\tilde{\gamma}_{12}(\omega_k) = \sin(\alpha_{12}(\omega_k))$.

Although the mixing matrix (10) satisfies our prerequisites the mixed signals are not always adequate from a perceptual point of view. Specifically, they are inadequate in case noise signals exhibit different short-term PSDs (as discussed in Sec. III B). Let us assume that two mutually independent babble speech signals are mixed using Eq. (10). The sensor signal $X_1(\ell,\omega_k)$ only consists of $N_1(\ell,\omega_k)$, while the sensor signal $X_2(\ell,\omega_k)$ consists of a mixture of $N_1(\ell,\omega_k)$ and $N_2(\ell,\omega_k)$. Since $N_2(\ell,\omega_k)$ is not present in $X_1(\ell,\omega_k)$ the sensors signals sound unnatural. To solve this problem we require that the contribution of the noise signals in each of the sensor signals is equal. In terms of the mixing matrix we require that $(C_{11})^2 = (C_{12})^2$ and $(C_{21})^2 = (C_{22})^2$, where C_{pq} denotes the element in the pth row and qth column of the matrix $\mathbf{C}$. Since the norm of the column vectors and the inner product between the column vectors are not affected by the rotation operation, a valid solution can be obtained by properly rotating the mixing matrix.

Define the rotation matrix as

$$\mathbf{R}(\omega_k) = \begin{bmatrix} \cos(\beta(\omega_k)) & -\sin(\beta(\omega_k)) \\ \sin(\beta(\omega_k)) & \cos(\beta(\omega_k)) \end{bmatrix}. \tag{11}$$

The rotated mixing matrix is then given by

$$\begin{aligned} \mathbf{C}'(\omega_k) &= \mathbf{R}(\omega_k)\mathbf{C}(\omega_k) \\ &= \begin{bmatrix} \cos(\beta(\omega_k)) & \sin(\alpha_{12}(\omega_k) - \beta(\omega_k)) \\ \sin(\beta(\omega_k)) & \cos(\alpha_{12}(\omega_k) - \beta(\omega_k)) \end{bmatrix}. \end{aligned} \tag{12}$$

The additional requirement is fulfilled when

$$(\cos(\beta(\omega_k)))^2 = (\sin(\alpha_{12}(\omega_k) - \beta(\omega_k)))^2. \quad (13)$$

Given $\alpha_{12}(\omega_k)$ we can easily solve $\beta(\omega_k)$, viz., $\beta(\omega_k) = \alpha_{12}(\omega_k)/2 - \pi/4 + \pi n$, where $n \in \mathbb{Z}$.

B. Multiple sensors

In this section we develop an efficient technique to generate multiple sensor signals with a predefined spatial coherence.

The first technique that could be used is based on simulating the actual physical properties of the noise field, as shown in Ref. 10. However, a large number of mutually independent noise sources is required to approximate the desired spatial coherence.[10] Furthermore, the obtained spatial coherence depends on the positions of the noise sources. Therefore it is difficult, if not impossible, to find the appropriate positions to simulate an arbitrary spatial coherence.

The second technique, related to Eq. (10), is based on the Cholesky decomposition.[13] This technique is used in econometrics,[14] and in communications[15] to generate random variables with specific statistical properties. Accordingly, we obtain the mixing matrix by computing the Cholesky decomposition of the matrix $\tilde{\mathbf{\Gamma}}(\omega_k)$, i.e.,

$$\tilde{\mathbf{\Gamma}}(\omega_k) = \mathbf{C}^H(\omega_k)\mathbf{C}(\omega_k), \quad (14)$$

where $\mathbf{C}(\omega_k)$ is an upper triangle matrix. Since the mixing matrix is an upper triangle matrix we encounter a similar problem as in the two sensor case. Here, the Mth sensor signal consists of all noise signals while the first sensor signal only consists of the first noise signal. Furthermore, it should be noted that the Cholesky decomposition can only be calculated in case the matrix $\tilde{\mathbf{\Gamma}}(\omega_k)$ is positive definite. This requirement is fulfilled when Eq. (5) or Eq. (6) is used. In addition, the condition number of $\tilde{\mathbf{\Gamma}}(\omega_k)$ should not be too close to zero. The latter occurs when the spatial coherence corresponds to a directional sound source. In that case all values are nonzero and the rank of $\tilde{\mathbf{\Gamma}}(\omega_k)$ equals to 1. Hence, the response to a directional source cannot be simulated using this technique.

A more general solution is obtained by calculating the eigenvalue decomposition (EVD) of the matrix $\tilde{\mathbf{\Gamma}}(\omega_k)$ as follows:

$$\tilde{\mathbf{\Gamma}}(\omega_k) = \mathbf{V}(\omega_k)\mathbf{D}(\omega_k)\mathbf{V}^H(\omega_k). \quad (15)$$

Now we can split the diagonal matrix $\mathbf{D}(\omega_k)$ to obtain

$$\tilde{\mathbf{\Gamma}}(\omega_k) = \mathbf{V}(\omega_k)\sqrt{\mathbf{D}(\omega_k)}\sqrt{\mathbf{D}(\omega_k)}\mathbf{V}^H(\omega_k). \quad (16)$$

The mixing matrix is then given by

$$\mathbf{C}(\omega_k) = \sqrt{\mathbf{D}(\omega_k)}\mathbf{V}^H(\omega_k). \quad (17)$$

For $M > 2$ the mixing matrix does not provide equal contribution of each of the noise signals in each of the sensor signals. However, the results of informal listening tests confirmed that the sensor signals generated using the mixing matrix (17) are perceptually satisfactory. It should be noted that the EVD can also be used to compute the response to a single directional noise source. In case the spatial coherence is related to a coherent sound field only one eigenvalue is larger than zero. Therefore, only the first row of the mixing matrix will contain elements larger than zero. Consequently, the sensor signals are all related to the first noise signal.

TABLE I. Summary of the developed algorithm that generates multisensor signals under a predefined spatial coherence constraint.

(1) Define a matrix $\tilde{\mathbf{\Gamma}}(\omega_k)$ for each ω_k that consists of the predefined spatial coherence values.
(2) Calculate the eigenvalue decomposition of the matrix $\tilde{\mathbf{\Gamma}}(\omega_k) = \mathbf{V}(\omega_k)\mathbf{D}(\omega_k)\mathbf{V}^H(\omega_k)$. The mixing matrix is then obtained by $\mathbf{C}(\omega_k) = \sqrt{\mathbf{D}(\omega_k)}\mathbf{V}^H(\omega_k)$.
(3) Generate M mutually independent complex random signals $N_p(\ell, \omega_k)$ (see Sec. III).
(4) For all ℓ and ω_k. Calculated $\mathbf{X}(\ell, \omega_k) = \mathbf{C}^H(\omega_k)\mathbf{N}(\ell, \omega_k)$, where $\mathbf{X}(\ell, \omega_k) = [X_1(\ell, \omega_k), \ldots, X_M(\ell, \omega_k)]^T$ and $\mathbf{N}(\ell, \omega_k) = [N_1(\ell, \omega_k), \ldots, N_M(\ell, \omega_k)]^T$.
(5) Finally, the sensor signals can be obtained by calculating the inverse STFT of $X_p(\ell, \omega_k)$ for $p \in \{1, 2, \ldots, M\}$.

V. ALGORITHM SUMMARY AND COMPUTATIONAL COMPLEXITY

A summary of the developed algorithm for generating M stationary or nonstationary sensor signals is provided in Table I. The first three steps are part of the initialization. The fourth step generates the STFT coefficients of the sensor signals. Finally, in the fifth step, the inverse STFT is used to obtain M discrete time signals that exhibit the predefined spatial constraint.

We now determine the computational complexity of the algorithm. Let us assume that the frame length of the STFT equals K. Firstly, we determine the computational complexity required to obtain the mixing matrix. Since the mixing matrix is time invariant we only need to compute it once. The complexity of constructing a mixing matrix for a single frequency is $O(M^3)$ for the EVD and $O(M^2)$ to construct the mixing matrix. For K frequencies the complexity is $O(K(M^2 + M^3))$. Hence, the computational complexity of the initialization grows rapidly for increasing K and M. Secondly, we determine the computational complexity required to compute Eq. (8). Specifically, we need to generate M STFT coefficients, one for each sensor signal, which yields $O(M^2)$ per time frame and frequency index. For K frequencies we then obtain $O(KM^2)$.

Let us assume that we need to generate M sensor signals of length L. In case $R = K/4$ denotes the number of samples between two successive STFT frames, we need to compute $L' = \lceil L/R \rceil$ time frames. The computational complexity of computing the STFT coefficients for all time frames, frequencies, and sensor signals yields $O(L'KM^2)$. Finally, we compute the inverse STFT, which can be efficiently implemented using a weighted overlap-add technique. The discrete Fourier transform of length K can be computed using the fast Fourier transform and has a complexity of order $O(K \log_2(K))$. The total computational complexity to compute $L \times M$ samples is given by $O(L'K^2M^2 \log_2(K)) \approx O(LKM^2 \log_2(K))$. We can see that the computational complexity per sample grows for increasing K and M. For a

given M the overall computational complexity can be reduced by using a small value of K. The minimum value of the frame length K is determined by the minimum filter order. The influence of K on the accuracy of the spatial coherence is examined in Sec. VI.

The computational complexity can be reduced by using more efficient techniques to compute the EVD, by performing more efficient matrix multiplications, and by exploiting the fact that the spectrum is conjugate symmetric.

VI. PERFORMANCE EVALUATION

In this section we analyze the generated sensor signals. The spatial coherences between two sensors in spherical and cylindrical noise fields are calculated and depicted. Firstly, the sensor signals received in a homogenous noise field are generated as described in Sec. III A and analyzed. Secondly, babble speech is generated as described in Sec. III B and subsequently analyzed. For the following analysis we have generated two sensor signals since the spatial coherence can only be measured between two sensors. The authors conducted the analysis with more than two sensors and confirmed that the results are equivalent.

In this work we used a recorded single babble speech signal available in Ref. 16. Alternatively, one can generate a babble speech signal by mixing N mutually independent speech signals. In order to obtain a realistic babble speech signal one can convolve each speech signal with a different room impulse response (the source positions are uniformly spread in a room). The N room impulse responses (RIRs) can be generated using the image method proposed by Allen and Berkley[17] with a reverberation time between 0.3 and 0.6 s. This method can also be used to generate M babble speech signals that exhibit the characteristics of a 2D or a 3D diffuse noise field. In the latter case we require MN RIRs. In case multiple sensor signals are generated, it is recommended to use only the reverberant parts of the RIRs since the direct paths are coherent. When multiple babble speech signals are generated the spatial coherence function is determined by the geometry of the enclosure and acoustic properties of the walls. Specifically, a spherical (3D diffuse) noise field is obtained when the reflection coefficients of the six walls are larger than zero. A cylindrical (2D) noise field is obtained when the reflection coefficients of the floor and ceiling are zero, and the other reflection coefficients are larger than zero.

Here we have used the later method for comparison, i.e., we generated multiple sensor signals by mixing 50 reverberant noise signals; this method is indicated as the "image method". The reverberant noise signals are generated by convolving each noise source signal with the reverberant part of a RIR. The noise sources are positioned such that (i) the distance between each source and all walls is larger than 1 m, and (ii) the distance between each source and the center of the sensor array is larger than 1 m. In order to simulate a spherical noise field the reflection coefficients of the six walls are set to 0.8. To simulate a cylindrical noise field the reflection coefficients of the sidewalls are set to 0.8 while those related to the floor and ceiling are set to 0. It should be noted that the computational complexity of the developed algorithm is much smaller than that of the image method. The developed algorithm only requires the generation of M noise sources, while the image method requires the generation of N noise sources, where $N \gg M$ to approximate the desired noise field. For N noise sources and M sensors we need to compute MN RIRs. Subsequently, each noise source needs to be convolved with M filters. Hence, the image method requires MN filter operations while the developed algorithm requires only M^2 filter operations.

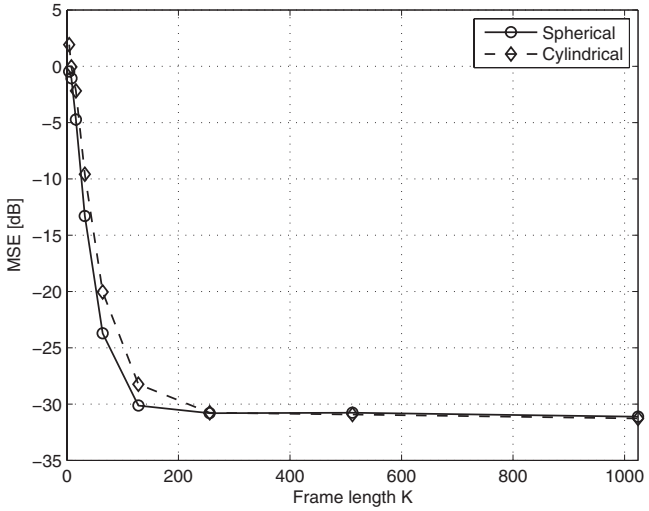


FIG. 1. Minimum square error between the theoretical spatial coherence and the spatial coherence of the generated signals versus the STFT frame length K.

Let us define the error between the spatial coherence of two generated signals and the theoretical spatial coherence by the normalized mean square error (MSE), i.e.,

$$\mathrm{MSE} = \frac{\sum_{k=0}^{K/2-1} (\hat{\gamma}_{pq}(\omega_k) - \tilde{\gamma}_{pq}(\omega_k))^2}{\sum_{k=0}^{K/2-1} (\tilde{\gamma}_{pq}(\omega_k))^2}, \tag{18}$$

where k denotes the discrete frequency index and $\hat{\gamma}_{\mathrm{pq}}(\omega_k)$ denotes the estimated spatial coherence of the desired spatial coherence $\tilde{\gamma}_{\mathrm{pq}}(\omega_k)$. The MSE is used to evaluate the image method and the developed algorithm. The coherence between the pth and the qth sensor was estimated using Welch's averaged periodogram method.[18] We used the fast Fourier transform of length 2048, a Hann window, and 75% overlap.

A. Minimum frame length

As discussed in Sec. V we prefer to use short STFT frames to achieve low computational complexity. In this section we determine the minimum length of the STFT frames, or in other words the minimum filter length. For the following experiment the mutually independent noise signals were generated as described in Sec. III A. Subsequently, we generated two sensor signals obtained in spherically and cylindrically isotropic noise fields as described in Sec. IV. The obtained MSE values for different values $K \in \{4, 8, 16, \ldots, 1024\}$ are shown in Fig. 1. For both noise fields we can see that the MSE is not significantly improved when $K \geq 256$. Hence, we chose $K = 256$, which provides a low MSE and results in a computationally efficient algo-

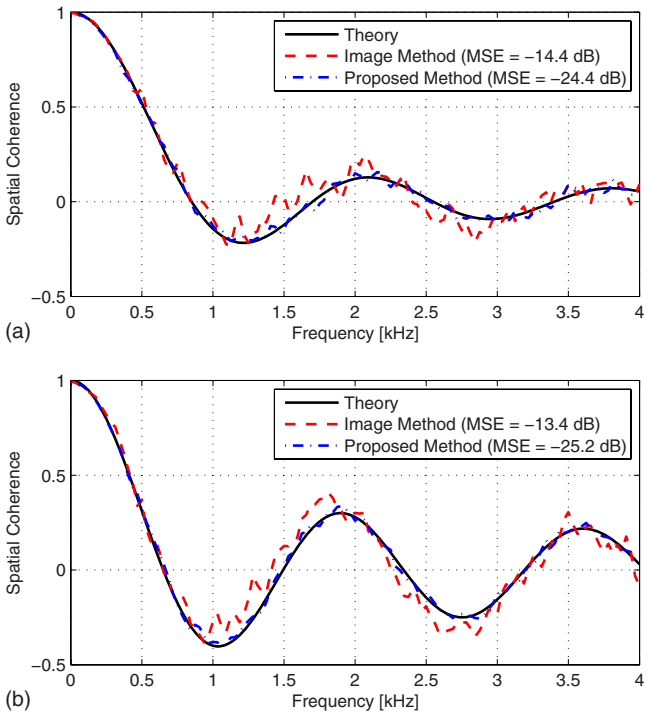


FIG. 2. (Color online) Spatial coherence between two nonstationary speechlike noise signals (d=20 cm): (a) spherically isotropic noise field and (b) cylindrically isotropic noise field.

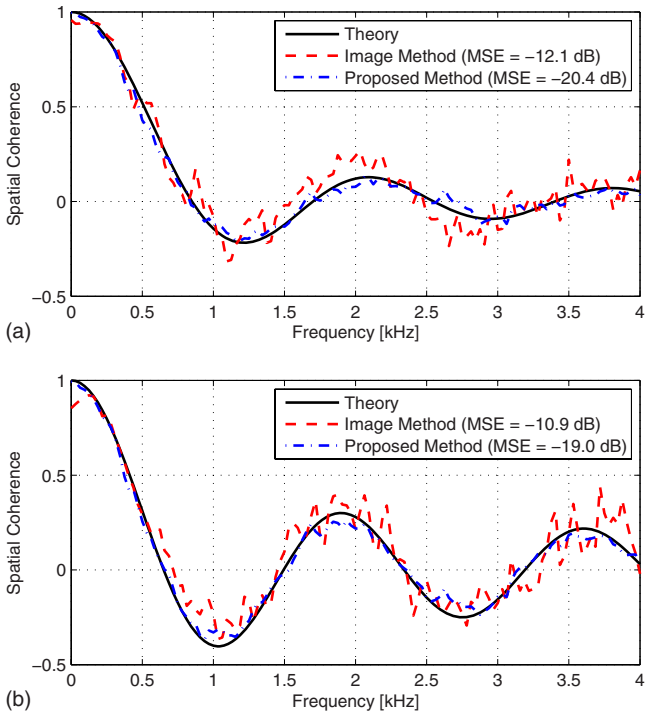


FIG. 3. (Color online) Spatial coherence between two babble speech sensor signals (d=20 cm): (a) spherical noise field and (b) cylindrical noise field.

rithm. Similar results were obtained when the noise signals are generated using the method described in Sec. III B.

B. Isotopic noise

The developed algorithm, as summarized in Table I, was used to generate the sensor signals obtained in spherically and cylindrically isotropic noise fields. Here we used nonstationary speechlike noise signals that were generated using the method described in Sec. III A. For comparison we also generated a set of sensor signals using the image method. We generated two sensor signal of 20 s and intersensor distance d=20 cm.

The simulation and theoretical results for the spherical and cylindrical noise fields are shown in Fig. 2. From the results shown in this figure we can see that the spatial coherence of the generated sensor signals obtained using the proposed method closely matches the theoretical spatial coherence for all frequencies. The spatial coherence of the generated sensor signals obtained using the image method only gives a good match at low frequencies. The MSE obtained by the proposed method is significantly lower than the MSE obtained by the image method for both noise fields.

C. Babble speech

Now the developed algorithm was used to generate the nonstationary sensor signals in spherically and cylindrically isotropic noise fields. The noise signals consist of babble speech signals that were generated using the method described in Sec. III B. We generated two sensor signals of 20 s and intersensor distance d=20 cm.

The simulation and theoretical results for the spherical and cylindrical noise fields are shown in Fig. 3. From these results we can see that the spatial coherence of the generated signals closely matches the theoretical spatial coherence. Again we see that the MSE of the proposed method is significantly smaller than the MSE of the image method. In Fig. 4 the long-term PSDs of sensor signals 1 and 2 are shown. We can see that the long-term PSDs of the multisensor babble speech signals are approximately equal although the short-term PSD of the babble speech signal at each time frame is known to be different. This demonstrates that we can accurately generate diffuse babble speech.

VII. CONCLUSION

We have developed a computationally efficient algorithm to generate sensor signals that exhibit a predefined spatial coherence. Different noise types such as babble

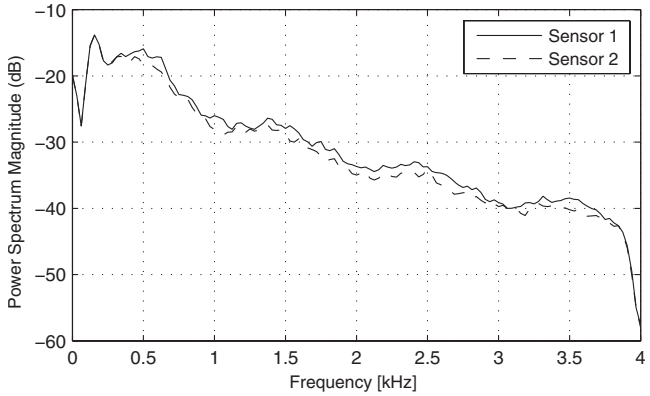


FIG. 4. Estimated PSD of sensor signals 1 and 2.

speech and factory noise can easily be generated. The algorithm consists of two steps: Firstly, we generate M mutually independent noise signals in the STFT domain, where M is equal to the number of sensors. Secondly, we calculate a frequency dependent mixing matrix, which will induce the predefined spatial coherence. We generate the spectral coefficients of the sensor signals for each time frame by multiplying the mixing matrix with the spectral coefficients of the noise signals. The sensor signals are then obtained by computing the inverse STFT. The major benefits of the developed algorithm are that (i) we can induce any spatial coherence and (ii) we only require M mutually independent noise signals. The performance evaluation showed that the spatial coherence of the generated sensor signals closely resembles the desired spatial coherence. Therefore the generated signals are useful for the evaluation and analysis of various signal processing algorithms. In addition, we showed that the MSE between the desired spatial coherence and the spatial coherence of generated signals is smaller than the MSE between the desired spatial coherence and the spatial coherence obtained using the image method that is frequently used in the acoustic signal processing community.

[1]G. Elko, "Superdirectional microphone arrays," in *Acoustic Signal Processing for Telecommunication*, edited by S. Gay and J. Benesty (Kluwer, Hingham, MA, 2000), Chap. 10, pp. 181–237.

[2]S. Gannot and I. Cohen, "Speech enhancement based on the general transfer function GSC and postfiltering," IEEE Trans. Speech Audio Process. **12**, 561–571 (2004).

[3]J. Bitzer, K. Simmer, and K. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'99), 1999, Vol. **5**, pp. 2965–2968.

[4]J. Bitzer, K. Simmer, and K. Kammeyer, "Multimicrophone noise reduction by postfilter and superdirective beamformer," in Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC'99), 1999, pp. 100–103.

[5]S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive array noise suppression of hands-free speaker input in cars," IEEE Trans. Veh. Technol. **42**, 514–518 (1993).

[6]N. Dal-Degan and C. Prati, "Acoustic noise analysis and speech enhancement techniques for mobile radio applications," Signal Process. **18**, 43–56 (1988).

[7]G. Elko, "Spatial coherence functions," in *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward (Springer, New York, 2001), Chap. 4, pp. 61–85.

[8]B. F. Cron and C. H. Sherman, "Spatial-correlation functions for various noise models," J. Acoust. Soc. Am. **34**, 1732–1736 (1962).

[9]M. Goulding and J. Bird, "Speech enhancement for mobile telephony," IEEE Trans. Veh. Technol. **39**, 316–326 (1990).

[10]E. Habets and S. Gannot, "Generating sensor signals in isotropic noise fields," J. Acoust. Soc. Am. **122**, 3464–3470 (2007).

[11]H. Cox, "Spatial correlation in arbitrary noise fields with application to ambient sea noise," J. Acoust. Soc. Am. **54**, 1289–1301 (1973).

[12]R. Cook, R. Waterhouse, R. Berendt, S. Edelman, and M. Thompson, "Measurement of correlation coefficients in reverberant sound fields," J. Acoust. Soc. Am. **27**, 1072–1077 (1955).

[13]E. Kreyszig, *Advanced Engineering Mathematics*, 8th ed. (Wiley, New York, 1998).

[14]R. Tsay, *Analysis of Financial Time Series*, Probability and Statistics (Wiley, New York, 2002).

[15]B. Natarajan, C. Nassar, and V. Chandrasekhar, "Generation of correlated Rayleigh fading envelopes for spread spectrum applications," IEEE Commun. Lett. **4**, 9–11 (2000).

[16]TNO-Perception, "Babble speech recording," http://spib.rice.edu/spib/data/signals/noise/babble.html (last viewed June, 2008).

[17]J. Allen and D. Berkley, "Image method for efficiently simulating small room acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).

[18]P. Stoica and R. Moses, *Spectral Analysis of Signals* (Prentice-Hall, Englewood Cliffs, NJ, 2005).

Membrane covered duct lining for high-frequency noise attenuation: Prediction using a Chebyshev collocation method

Lixi Huang[a)]
Department of Mechanical Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China

(Received 13 December 2007; revised 26 June 2008; accepted 30 July 2008)

A spectral method of Chebyshev collocation with domain decomposition is introduced for linear interaction between sound and structure in a duct lined with flexible walls backed by cavities with or without a porous material. The spectral convergence is validated by a one-dimensional problem with a closed-form analytical solution, and is then extended to the two-dimensional configuration and compared favorably against a previous method based on the Fourier–Galerkin procedure and a finite element modeling. The nonlocal, exact Dirichlet-to-Neumann boundary condition is embedded in the domain decomposition scheme without imposing extra computational burden. The scheme is applied to the problem of high-frequency sound absorption by duct lining, which is normally ineffective when the wavelength is comparable with or shorter than the duct height. When a tensioned membrane covers the lining, however, it scatters the incident plane wave into higher-order modes, which then penetrate the duct lining more easily and get dissipated. For the frequency range of $f=0.3-3$ studied here, $f=0.5$ being the first cut-on frequency of the central duct, the membrane cover is found to offer an additional 0.9 dB attenuation per unit axial distance equal to half of the duct height. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977743]

PACS number(s): 43.50.Gf, 43.20.Tb, 43.20.Mv, 43.20.Bi [KA] Pages: 2918–2929

I. INTRODUCTION

Duct noise control finds numerous domestic and industrial applications, such as ventilation pipework, vehicle exhaust, etc. Duct lining is perhaps the most reliable and mature technique. There have been some hygienic concerns over the use of certain fibers, but the use of microperforated panels, and the potential development in metallic foams may overcome such drawbacks. In this study, we are primarily interested in addressing the acoustic issue of high-frequency noise absorption instead of the practical application issues. Due to the grazing incidence, sound absorption by a lining is not as easy as for normal or random incidence encountered in a room or outdoor environment. The performance of duct lining is characterized by the normalized attenuation, conventionally denoted as L_h, and defined as the reduction in sound pressure level (SPL) per unit axial distance equal to half of the duct opening dimension, which is denoted as $h/2$ in this study (Fig. 1). For ducts of rectangular cross section, L_h reaches its peak value of around 3–4 dB in the medium-frequency range of $f=0.5-1.0$, where f is a dimensionless frequency defined as $f=f^*h/c_0$, f^* is the dimensional frequency, c_0 is the speed of sound in air, and h is the full duct height. In other words, the peak frequency is such that the wavelength is between the half and the full opening dimension. The precise peak value and the frequency location in an attenuation spectrum depend also on the lining depth and on the porosity (or flow resistivity); see, for example, Fig. 9.22 of Munjal *et al.* (2006). According to the curves shown in this figure for a lining depth equal to the half-gap, the attenuation rate L_h drops to around 0.3 dB for $f=3$ and the effect of porosity does not have a major impact in this particular case. Note that the attenuation rate quoted here is for the plane incident wave, which is the least attenuated mode.

Part of the reason for such a reduced performance at high frequencies is that there is simply not much chance for the sound waves to penetrate into the lining. If the incident sound propagates in higher-order modes with cross sectional air particle oscillations, the penetration into and absorption by the lining would be much improved. Based on such expectation, this study considers the configuration of duct lining with an impervious membrane cover. The use of a membrane cover has been traditionally associated with the protection of the porous lining from the abrasive flow carried in the duct and its acoustic effect is normally counterproductive (Munjal and Thawani, 1997) except around a low-frequency resonance it may introduce (Horoshenkov and Sakagami, 2001), in which case the membrane, or a plate, mainly serves as an additional mass of the system leading to a reduced system resonance frequency. In this study, however, the membrane is stretched under tension and its main physical ingredient is its tensile restoring force. The incident wave excites the membrane, which in turn scatters the incident wave. It is hoped that the induced membrane vibration channels the incident wave into the porous layer underneath the membrane.

The basic configuration is shown in Fig. 1, which is also used to illustrate the numerical modeling in the next section. The device is essentially an expansion chamber with each side-branch cavity filled by a porous material and covered by a tensioned membrane. Without the porous material, the

a)Author to whom correspondence should be addressed. Electronic mail: lixi.huang@hku.hk

 © 2008 Acoustical Society of America

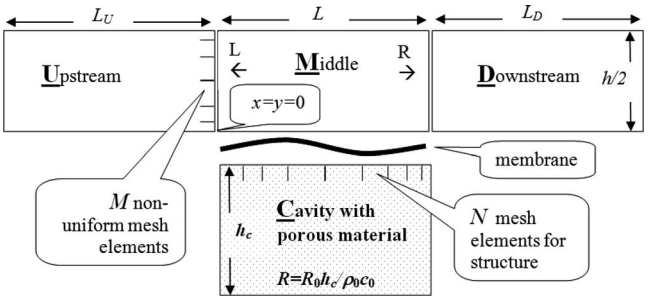


FIG. 1. Geometry and domain decomposition for the ChC method. The duct height is h and only the lower symmetrical half is solved. The cavity is covered by a membrane and is filled with a sound absorption material of dimensional flow resistivity R_0.

membrane vibration has been shown to reflect the grazing incident sound over a broad frequency band (Huang, 1999, 2002). For optimal performance in the very low-frequency region, the tension required is very high and thus it has been called a drumlike silencer (Huang, 2002; Choy and Huang, 2002). With the absorption material, the device becomes mainly dissipative and, as shown in this study, the contribution of sound reflection is minimal in the new configuration. The effect of a mean flow is expected to be significant only when the Mach number is large enough, and is therefore not considered here. Besides, when an impervious membrane covers the duct lining, there would be no flow entering into the porous material. The issue of potential flow induced instability of similar configuration has been studied elsewhere (Choy *et al.* 2007).

Apart from exploring the physics of the high-frequency sound propagation in a membrane covered duct lining, this study is also motivated by the testing of a numerical method for the full coupling between sound and vibration. When the lining is infinitely long, analytical modeling, e.g., (Ingard, 1994), can be employed to find the complex wavenumber along the duct, hence the normalized attenuation rate. For a finite segment of lining, the scattering by the two junctions is important, and the performance should be predicted by treating the lining as bulk reactive. For very low frequencies, plane waves can be assumed to obtain approximate solution for practical design purposes. For higher frequencies and slightly more complex geometries, such as an expansion chamber with inlet and exit tube extensions, or split silencers, mode-matching has been an important semianalytical tool, e.g., (Mechel, 1990) and (Cummings, 1999), for duct lined with a locally reactive porous material. The difficulty then is the finding of all complex eigenvalues for the lined passages. To overcome such difficulty, a root-free method of mode-matching has also been put forward (Lawrie and Kirby, 2006). A cousin of such mode-matching method is the collocation method, which matches the values of pressure and particle velocity at individual points along the interface. Kirby and Lawrie (2005) introduced this method for the problem of sound attenuation in a duct lined with a bulk reactive porous material and it can be extended to a fairly general type of interface conditions. The relationship between the acoustic mode-matching and its point-collocation cousin mirrors that between spectral (or tau) and pseudospectral (or collocation) methods to be described below. Broadly speaking, the matching of expansion coefficients and the matching of values at grid nodes should produce the same answer. The major difference between the acoustic mode-matching discussed here and the general spectral methods is the following. The acoustic mode-matching expands the unknown in terms of an orthogonal set of eigenfunctions for the specific problem at hand, while the spectral methods seek to interpolate the unknown in terms of the general functions such as Fourier or Chebyshev series. In spectral methods, there is no need to solve the eigenvalue problem since the boundary condition is only satisfied by the sum of the truncated series. The use of Chebyshev series carries extra benefit since it can accommodate arbitrary boundary conditions. On the other hand, the labor in finding the eigenvalues or eigenfunctions in the acoustic mode-matching may be rewarded by fast computational speed at its later stage. However, the need to tackle complex geometry at medium to high frequencies still calls for full numerical solutions such as the finite element method (FEM), e.g., (Peat and Rathi, 1995), or boundary element method (BEM), e.g., (Cheng *et al.* 1991). Although certainly suitable for the current configuration, the BEM is less generally applicable than the FEM and is not considered further in the current discussion.

In the current study, the spectral method of Chebyshev collocation (ChC) is introduced together with the technique of domain decomposition. Despite being a weighted residual method, a spectral collocation method, namely, the pseudospectral method, is closer in implementation to the finite difference method than to the FEM. Therefore, the results from this proposed method are first compared with a fourth-order finite difference method (FD4) for a simple problem with a closed-form analytical solution. Then it shall be compared with FEM and another spectral method used in our previous study on similar configurations (Huang, 1999). The rationale for introducing the new method is described as follows. When new physics is the focus of a problem, it is desirable to have high accuracy in the numerical solution so that numerical error accumulation does not hinder the interpretation of small quantities, such as the transmitted sound wave through a very effective silencer. There are two approaches to obtaining such high accuracy. One is through the refinement of mesh and the other is to use schemes of high order in discretization (Boyd, 2001). For very smooth problems, such as the sound-vibration coupling problem being tackled here, the high-order approach is suitable. It can be either implemented in the FEM within individual elements or by the spectral method globally. The latter tends to produce higher accuracy at the expense of having full coefficient matrices as opposed to sparse or banded matrices in FEM. In other words, when the level of accuracy is fixed, the required mesh density in the spectral method can be lower than that of the FEM, but the computational time may grow rapidly with the problem size. This is demonstrated in Fig. 6 when the accuracy and CPU time are compared between the present method and the FEM.

Spectral methods can be categorized by the basis functions used. Fourier methods are suitable for space periodic problems, while Chebyshev or Legendre functions are better

for general types of problems (Gottlieb *et al.* 1984). The Chebyshev method is chosen for the current study. The method adopted in the previous studies for this particular type of sound-structure interaction problems (Huang, 1999) is along the lines of the Galerkin procedure and is equivalent to the spectral method based on the Fourier series. The use of the Fourier based methods for nonperiodic problems causes Gibbs phenomenon at the boundaries. One of the objectives of the current study is to investigate the precise nature of this Gibbs phenomenon and the ease with which the problem can be rectified by the Chebyshev functions. When coupled with domain decomposition, the current method becomes a prototype of the so-called spectral element method (Patera, 1984; Karniadakis and Sherwin, 2005), which now holds the promise of solving real-world problems in a way the conventional FEM does.

In what follows, Sec. II is devoted to the validation of the solution method and Sec. III focuses on issues relevant to the physics of high-frequency sound absorption by duct lining. In Sec. II, the validation begins with a one-dimensional, frequency-domain problem of sound reflection and transmission through a segment of different wave media. Part of the reason for considering such a simple model is to introduce the spectral method in its simplest setting. The other reason is that a closed-form exact solution exists. Having achieved the validation of the method in one dimension, Sec. II B introduces the current method for the two-dimensional problem of drumlike silencer (Huang, 2002) for frequencies below the first cut-on of the central duct. The focus in this subsection is the patching condition across the domains. In this study, the membranes have no thickness and are treated as domain interfaces. The dynamics of the tensioned membrane is enforced as a patching condition on the domain interface. In Sec. II C, detailed comparison is made with the previous method based on the Fourier–Galerkin (FGa) approach and the standard FEM. Section III begins with the implementation of the Dirichlet-to-Neumann (DtN) outgoing wave condition necessary for the propagation of high-order duct acoustics modes. In Sec. III B, the new method is used to predict the optimal parameter setting for the membrane covered duct lining, and the results are analyzed in Sec. III C. Conclusions are drawn in Sec. IV.

For simplicity, the following abbreviations are introduced in addition to FEM and DtN just mentioned.

- ChC: the spectral method of ChC with the Gauss–Lobatto grid.
- FGa: the spectral method of Galerkin procedure using Fourier series as the expansion basis with a uniform grid (Huang 1999).
- FD4: the fourth-order finite difference method following (Singer and Turkel, 1998).

II. NUMERICAL METHOD AND ITS VALIDATION

A. Validation in one dimension and comparison with a finite difference method

For the purpose of validating the spectral method by analytical solution, consider sound reflection in an otherwise uniform channel sandwiched by a segment with different normalized acoustic impedance, $Z=(\rho c)/(\rho_0 c_0)\neq 1$, from $x=0$ to $x=L$, where ρ and c are the fluid density and speed of sound, respectively, and subscript 0 denotes the region outside the sandwiched segment, or the ambient fluid. The governing Helmholtz equation follows from the wave equation in the frequency domain,

$$c^{-2}\partial^2 p/\partial t^2 - \nabla^2 p = 0 \rightarrow (\partial_x^2 + k^2)p = 0, \tag{1}$$

where p is the sound pressure, ω is the angular frequency, and $k=\omega/c$ is the wavenumber. The incident wave, denoted by subscript "in," is given as

$$p_{\text{in}} = 1 \times \exp(\imath\omega t - \imath k_0 x) = \rho_0 c_0 u_{\text{in}}, \tag{2}$$

where u is the acoustic particle velocity and $\imath=\sqrt{-1}$ is the unit imaginary number. The outgoing traveling wave condition for the reflected wave at the left-hand side, $x_{\text{left}}=-L_U$, is

$$p_{\text{ref}} = p - p_{\text{in}} = -\rho_0 c_0 (u - u_{\text{in}}) \rightarrow [\imath k p - \partial p/\partial x]_{x=x_{\text{left}}}$$
$$= -2\exp(-\imath k_0 x_{\text{left}}), \tag{3a}$$

and the anechoic condition at the right-hand side, $x_{\text{right}}=L+L_D$, is

$$p_T = \rho_0 c_0 u_T \rightarrow [\imath k p + \partial p/\partial x]_{x=x_{\text{right}}} = 0, \tag{3b}$$

where subscript T denotes the transmitted wave. The condition at the interface between two different acoustic media is the continuity of pressure p and the acoustic particle velocity u, which is obtained by the momentum equation $u=-(\rho\imath\omega)^{-1}\partial_x p$. The analytical solution for the amplitude of the transmitted wave is

$$|p_T| = 1/\sqrt{1 + \left[\tfrac{1}{2}(Z - Z^{-1})\sin(kL)\right]^2}. \tag{4}$$

In what follows, a FD4 is described first, followed by the ChC method. The FD4 construction follows (Singer and Turkel, 1998) by utilizing the Helmholtz equation to reduce the higher-order derivatives needed, $d^2p/dx^2=-k^2p$. Using a grid of uniform interval, Δx, the central finite difference scheme for the Helmholtz equation, Eq. (1), gives the following discretized governing equation for all interior grid points:

$$p_{j-1} - 2Ap_j + p_{j+1} = O(\Delta x^4),$$

$$A = \left\{1 - \tfrac{1}{2}(k\Delta x)^2\left[1 - \tfrac{1}{12}(k\Delta x)^2\right]\right\},$$

where coefficient A would have been unity if the second-order accuracy were acceptable. To be consistent, a fourth-order boundary condition is also derived. This essentially requires a single-sided first-order derivative. Denote the boundary index as 0 and its left- and right-hand sides as -1 and $+1$, respectively,

$$\partial_x p_0 = \frac{p_{+1} - Ap_0}{B\Delta x} + O(\Delta x^4) = \frac{Ap_0 - p_{-1}}{B\Delta x} + O(\Delta x^4),$$

$$B = 1 - (k\Delta x)^2/6.$$

When discretized, Eqs. (1), (3a), and (3b) become a set of algebraic equations with a matrix of three diagonals, which is easily solved.

For the ChC method, the Chebyshev polynomial of the first kind is used, $T_n(\tilde{x})=\cos(n\cos^{-1}\tilde{x})$, where $\tilde{x}\in[-1,+1]$ is the canonical coordinate signified by $\sim$, to which a practical coordinate, such as $x\in[0,L]$, is easily converted, as shown below. When the discretized representation is evaluated at the Gauss–Lobatto points,

$$x_i=(1+\tilde{x}_i)L/2,\quad \tilde{x}_i=-\cos[i\pi/N],\quad i=0,1,2,\ldots,N, \tag{5}$$

the nth derivative, $f^{(n)}=d^nf/dx^n$, at these $N+1$ points can be calculated by the product of a derivative matrix, denoted by square brackets, $[d^{(n)}]$, and the column vector of the function f evaluated at the $N+1$ grid points, $[f]\equiv[f(x_0)f(x_1)\cdots f(x_N)]^T$, where the superscript T denotes matrix transpose,

$$f^{(n)}(x_i)=\sum_{j=0}^{N} d_{i,j}^{(n)}f(x_j),\quad [f^{(n)}]=[d^{(1)}]^n[f],\quad d_{i,j}^{(1)}=\frac{2}{L}\tilde{d}_{i,j}^{(1)}, \tag{6a}$$

where the entries of the derivative matrix in the canonical form are given as follows:

$$\tilde{d}_{i,j}^{(1)}=\begin{cases}(1+2N^2)/6, & i=j=0\\ -\tilde{x}_i/[2(1-\tilde{x}_i^2)], & 0<i=j<N\\ (-1)^{i+j}(\bar{c}_i/\bar{c}_j)/(\tilde{x}_i-\tilde{x}_j), & i\neq j\\ -(1+2N^2)/6, & i=j=N,\end{cases}$$

$$\bar{c}_j=2,\quad j=0,N;\quad =1,\quad \text{otherwise}. \tag{6b}$$

Higher-order derivatives can be derived from the product of lower-order derivative matrices, e.g., $[\tilde{d}^{(2)}]=[\tilde{d}^{(1)}]^2$. However, direct assignment of entries for $[\tilde{d}^{(2)}]$, which is not listed here, can reduce round-off errors (Kosloff and Tal-Ezer, 1993; Bayliss *et al.*, 1994). Numerical integrations are conducted by the formula of Clenshaw and Curtis (1960) for even as follows:

$$\int_0^L f(x)dx\approx\frac{L}{2}\sum_{i=0}^{N}\varpi_i f(x_i)=\frac{L}{2}[\varpi][f], \tag{7a}$$

where $[\varpi]$ contains the weighting coefficients in the form of a row vector, whose entries are

$$\varpi_i=\begin{cases}1/(N^2-1), & i=0,N\\ \dfrac{4}{N}\displaystyle\sum_{j=0}^{N/2}\frac{1}{\bar{c}_j}\frac{\cos(2\pi ij/N)}{1-4j^2}, & \text{otherwise}.\end{cases} \tag{7b}$$

The governing Helmholtz equation, Eq. (1), is simply discretized as

$$\{[d_{i,j}^{(2)}]+k^2[I]\}[p]=0, \tag{8}$$

where $[I]$ stands for the unit diagonal matrix of a suitable size and $[p]=[p_0,p_1,\ldots,p_N]^T$ is the column vector of the sound pressure at the Gauss–Lobatto points. Equation (8) is, in fact, only applied at all interior points, $i=1,2,\ldots,N-1$, while the boundary conditions are enforced at the two ends, $i=0$, N, which are discretized for the general Robin type as follows:

TABLE I. List of solutions required in the domain decomposition method.

i.d.	Domain	Left boundary	Right boundary	Domain $p=$
I	Upstream	Eq. (2)	$p=0$	$p^{(I)}+\xi p^{(U)}$
U		$\imath k_0 p-\partial_x p=0$	$p=1$	
L	Middle	$p=1$	$p=0$	$\xi p^{(L)}+\chi p^{(R)}$
R		$p=0$	$p=1$	
D	Downstream	$p=1$	Eqs. (3a) and (3b)	$\chi p^{(D)}$

$$\partial_x p+qp=g\rightarrow\sum_{j=0}^{N}d_{i,j}^{(1)}p_j+qp_i=g,\quad i=0,N. \tag{9}$$

The two boundary equations replace the first and last rows of Eq. (8). The solution for the set of $N+1$ linear equations can be obtained by the method of eigenvalue decomposition (Lynch *et al.*, 1964). This method is expensive for one-off matrix inversion but becomes economical when used repeatedly, as is the case in the two-dimensional case in the next section. For any nonsingular matrix D,

$$D=P\Lambda P^{-1},\quad D^n=P\Lambda^nP^{-1}, \tag{10}$$

where Λ is the diagonal matrix whose entries are the eigenvalues of D, and P is the matrix whose columns are the corresponding eigenvectors. Matrix inversion and square-root operations are computed by $n=-1$, $\frac{1}{2}$, respectively. The standard routine of "eig" in MATLAB® version 7 is used in the current study to perform the task of eigenvalue decomposition.

For the problem of sound propagation through a segment of different acoustic media, which resembles Fig. 1 with the cavity ignored, the physical domain is divided into three subdomains, which are solved separately by the procedure of domain decomposition (Peyret, 2002). It essentially involves assigning homogeneous or unit values for the unknown interfaces, as listed in Table I for the current one-dimensional problem. The first column assigns the "i.d." for each subdomain problem, which shall be used as superscripts to identify their solutions. The upstream and middle domains are solved twice with different boundary conditions, as specified by the third and fourth columns of Table I. The reconstructed solution for each subdomain is given in the last column, where variables ξ and χ are the unknown pressures at the two interfaces, $\xi=p|_{x=0}$, $\chi=p|_{x=L}$, to be determined by the velocity continuity at the interfaces,

$$u_{x=0-}^{(I)}+\xi u_{x=0-}^{(U)}=\xi u_{x=0+}^{(L)}+\chi u_{x=0+}^{(R)},$$

$$\xi u_{x=L-}^{(L)}+\chi u_{x=L-}^{(R)}=\chi u_{x=L+}^{(D)}, \tag{11}$$

where u is obtained from each subdomain solution via the momentum equation $u=-(\rho\imath\omega)^{-1}[d^{(1)}][p]$, and $u_{x=0-}^{(I)}$ means the velocity at the left-hand side of interface $x=0$ in problem I. Note that problem (I) accounts for the incident wave effect with a homogeneous right boundary condition, $p=0$, and that the homogeneous version of the inlet boundary condition, Eq. (3a), is used in problem (U) so that the full pressure satisfies the inlet boundary condition. The transmitted wave

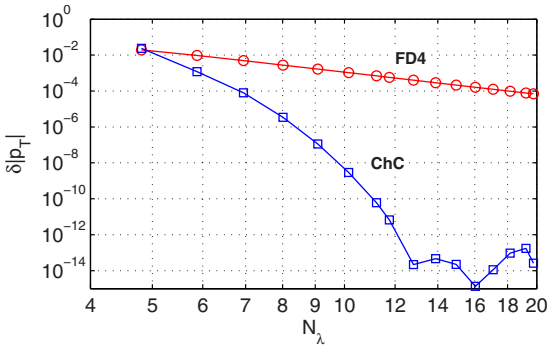


FIG. 2. (Color online) Comparison of the error for the transmitted wave between the spectral method of ChC and the FD4. The sandwiched segment has $\rho=\rho_0$, $c=3c_0$, and $L=5.615\lambda_0=1.87\lambda$. $N_\lambda=\lambda/\Delta x$ is the number of grids per wavelength.

is calculated as $p_T=\chi p^{(D)}_{x=L+L_D}$, and the magnitude $|p_T|$ can be compared directly with the exact solution (4).

The error is plotted in Fig. 2 as a function of the number of grids per wavelength, $N_\lambda=\lambda/\Delta x$, which is kept the same for all subdomains in this problem. Coincidentally, the two methods perform about the same for the lowest mesh density used here, but ChC performs far better as N_λ increases. The FD4 error has a constant logarithmic slope of $d(\log \text{Error})/d(\log \Delta x)\approx 4$. In other words, Error=const $\times\Delta x^4$, as expected. The ChC curve, on the other hand, reaches its bottom around $N_\lambda=13$ and further increase in mesh density leads to random variations associated with the round-off errors in the 32 bit floating point computation. The curve before reaching this limit approximately follows $\delta|p_T|=\exp(-0.1025(\lambda_0/\Delta x)^{2.2})$, confirming the exponential convergence for the spectral method.

B. Modeling of sound-membrane interaction

As shown in Fig. 1, the two-dimensional problem is divided into four subdomains and the tensioned membrane is the interface between the middle and the cavity domains. For the middle domain, it is discretized by $N+1$ Gauss–Lobatto points in the x direction and $M+1$ points in the y direction. The pressure over this domain is represented by a matrix $[p]$ of size $(M+1)$ by $(N+1)$, which is essentially a transposed image of the physical domain. Derivatives in x and y directions operate on its columns and rows, respectively. The discretization of the Helmholtz equation, Eq. (1), becomes

$$[d^{(xx)}][p]+[p][d^{(yy)}]^T+k^2[p]=[0], \tag{12}$$

where $d^{(xx)}$ and $d^{(yy)}$ are the second derivative matrices with sizes of $(N+1)\times(N+1)$ and $(M+1)\times(M+1)$, respectively. However, the equations should be replaced with appropriate boundary conditions on all boundary grids, details of which are given later. For example, the first row of $[d^{(xx)}]$ in the above equation will be replaced with the first row of $[d^{(x)}]$ when the boundary condition involving $\partial p/\partial x$ on the left-hand side is enforced. The incident wave condition given in Eq. (3a) is inhomogeneous, so is the complete discretized system, which is symbolically written as

$$D^{(x)}[p]+[p][D^{(y)}]^T+k^2[p]=[H], \tag{13}$$

where $D^{(x)}$ and $D^{(y)}$ represent, respectively, all derivative operators in x and y directions with boundary conditions implemented. This matrix equation has the form of $AX+XB=C$ and the procedure of Bartels and Stewart (1972) can be used. Using the matrix diagonalization technique expressed in Eq. (10),

$$D^{(x)}=\mathcal{A}S_x\mathcal{A}^{-1},\quad D^{(y)}=\mathcal{B}S_y\mathcal{B}^{-1}.$$

Equation (13) can be solved by left multiplying $\mathcal{A}^{-1}$ and right multiplying $(\mathcal{B}^T)^{-1}$,

$$S_x[P]+[P]S_y+k^2[P]=[\mathcal{H}],$$

where

$$[P]=\mathcal{A}^{-1}[p](\mathcal{B}^T)^{-1},\quad [\mathcal{H}]=\mathcal{A}^{-1}[H](\mathcal{B}^T)^{-1}.$$

Since S_x and S_y are diagonal matrices with eigenvalues, denoted by $\lambda_{x,i}$, $i=0,1,2,\ldots,N$ and $\lambda_{y,j}$, $j=0,1,2,\ldots,M$, respectively, the solution for the whole pressure matrix, $[p]$, is found by the following two steps:

$$P_{i,j}=\frac{\mathcal{H}_{i,j}}{\lambda_{x,i}+\lambda_{y,j}+k^2},\quad [p]=\mathcal{A}[P]\mathcal{B}^T.$$

The above procedure is applicable to a general rectangular domain with well prescribed boundary conditions.

The problem of drumlike silencer (Huang, 2002) is now solved by the same procedure of domain decomposition as described in the one-dimensional example in the last subsection. The geometry and domain decomposition are illustrated in Fig. 1. In all examples given, two membranes are used in the configuration, and Fig. 1 shows one symmetrical half with half duct height $h/2$. The cavity depth $h_c=h$ is used in all numerical examples without sound absorption material, so that direct comparison with previous results in (Huang, 2002) can be made, but $h_c=h/2$ is used for duct lining in the next section. Hard wall boundary conditions are applied on the symmetrical line of $y=h/2$ and on all structural hard walls. The four subdomains are denoted by U, D, M, and C, representing "upstream," "downstream," "middle," and "cavity," respectively. For the middle domain, the solution is divided into three elementary solutions in which an inhomogeneous boundary condition (forcing term) is applied on one interface at a time. The three interfaces are the left (L), right (R), and membrane (M) interfaces. All the subdomain solutions listed in Table I are repeated here in two dimensions with necessary changes. For example, subdomain with the i.d. of L in Table I is the solution for the middle domain with the left interface being assigned the inhomogeneous condition of $p=1$ while the other boundaries remain homogeneous. The effect of this inhomogeneous boundary condition may be called a "point forcing" in one-dimensional case. In the two-dimensional case, such "forcing" is conducted for the $(M-1)$ interior grid points, $i=1,2,3,\ldots,M-1$, one at a time, and the value of particle velocity u on the right interface of $x=L-$ is a column vector of $(M-1)$ elements. The collection of results from the $(M-1)$ point forcings becomes an influence matrix of size $(M-1)\times(M-1)$, denoted by curly brackets $\{u^{(L)}_{x=L-}\}$. The product of this matrix with the

column of the unknown pressure on the left interface, $[\xi]=[\xi_1\xi_2\cdots\xi_{M-1}]^T$, becomes the particle velocity u at the right interface, $\{u^{(L)}_{x=L-}\}[\xi]$, contributed to by the forcing on the left interface. Such boundary inhomogeneity may be called "line forcing." Note that the membrane surface is treated as a solid wall in line forcing problems denoted by L and R.

In addition to the two line forcing calculations described above, the membrane vibration velocity, $\partial\eta/\partial t$, is unknown and has to be found by the same point and line forcing techniques in a fully coupled solution. The kinematic boundary condition on the membrane is the continuity of its normal velocity with that in the fluid, $v=\partial\phi/\partial y=\partial\eta/\partial t$, in which the vertical particle velocity v can be derived from the sound pressure gradient via the vertical momentum equation, $\rho_0\partial v/\partial t+\partial p/\partial y=0$. Finally,

$$\rho_0\iota\omega\frac{\partial\eta}{\partial t}+\frac{\partial p}{\partial y}\bigg|_{y=0+}=0, \tag{14a}$$

where density ρ_0 is replaced with the complex density of the sound absorption material when the same condition is applied on the lower side of the membrane facing the dissipative cavity, $y=0-$. In the ChC solution scheme, the membrane vibration velocity is the major unknown to be determined, and is denoted below by μ. The discrete form of Eq. (14a) is given as follows:

$$\mu_j=\frac{\partial\eta}{\partial t}\bigg|_{x_j}=\frac{-1}{\rho_0\iota\omega}\times\frac{\partial p(x_j,y=0+)}{\partial y},$$

$$j=0,1,2,\ldots,N. \tag{14b}$$

The continuity of u at the two interfaces is expressed as follows:

$$\text{Left}(x=0):\{u^{(I)}_{x=0-}\}+\{u^{(U)}_{x=0-}\}[\xi]=\{u^{(L)}_{x=0+}\}[\xi]+\{u^{(R)}_{x=0+}\}[\chi]+\{u^{(M)}_{x=0+}\}[\mu],$$

$$\text{Right}(x=L):\{u^{(D)}_{x=L+}\}[\chi]=\{u^{(L)}_{x=L-}\}[\xi]+\{u^{(R)}_{x=L-}\}[\chi]+\{u^{(M)}_{x=L-}\}[\mu]. \tag{15}$$

The use of Eqs. (14a) and (14b) has already satisfied the kinematic boundary condition for the structural surface. The structural dynamics serves as the matching condition between the middle and cavity domains,

$$m_0\frac{\partial^2\eta}{\partial t^2}-T_0\frac{\partial^2\eta}{\partial x^2}+p_{y=0+}-p_{y=0-}=0, \tag{16}$$

where m_0 and T_0 are, respectively, the membrane surface density and tensile force, respectively, $\eta=\mu/\iota\omega$ is the membrane displacement, and $p_{y=0\pm}$ are the fluid loading on the upper and lower sides of the structure, respectively, and are found by subdomain solutions as follows:

$$p_{y=0+}=\{p^{(L)}_{y=0+}\}[\xi]+\{p^{(R)}_{y=0+}\}[\chi]+\{p^{(M)}_{y=0+}\}[\mu],$$

$$p_{y=0-}=\{p^{(C)}_{y=0-}\}[\mu],$$

which can be inserted to Eq. (16) to give the following matching conditions on the structural surface $y=0$:

$$\iota\omega m_0[\mu]+\iota T_0\omega^{-1}[d^{(xx)}][\mu]+\{p^{(L)}_{y=0+}\}[\xi]+\{p^{(R)}_{y=0+}\}[\chi]+\{p^{(M)}_{y=0+}-p^{(C)}_{y=0-}\}[\mu]=[0]. \tag{17}$$

This equation is solved together with Eq. (15), which is collected together as follows:

$$\begin{bmatrix}(u^{(L)}_{x=0+}-u^{(U)}_{x=0-}) & u^{(R)}_{x=0+} & u^{(M)}_{x=0+}\\ u^{(L)}_{x=L-} & (u^{(R)}_{x=L-}-u^{(D)}_{x=L+}) & u^{(M)}_{x=L-}\\ p^{(L)}_{y=0+} & p^{(R)}_{y=0+} & (p^{(M)}_{y=0+}-p^{(C)}_{y=0-}+\mathcal{M})\end{bmatrix}\times\begin{bmatrix}\xi\\ \chi\\ \mu\end{bmatrix}=\begin{bmatrix}u^{(I)}_{x=0-}\\ 0\\ 0\end{bmatrix}, \tag{18a}$$

where

$$\mathcal{M}=m_0\iota\omega[I]+\iota T_0\omega^{-1}[d^{(xx)}] \tag{18b}$$

is a structural mechanics matrix. The transmitted wave at the downstream exit is given as follows together with the transmission loss (TL):

$$p_T=\frac{2}{h}\int_0^{h/2}p_{x=L+L_D}dy=[\varpi]\{p^{(D)}_{x=L+L_D}\}[\chi],$$

$$\text{TL}=-20\log_{10}|p_T|, \tag{19}$$

where $[\varpi]$ is the row vector from the Clenshaw–Curtis integration, Eqs. (7a) and (7b).

C. Comparison with FEM and Fourier spectral method

In order to assess the accuracy of the ChC method with domain decomposition, the results are compared with standard FEM and the FGa procedure used earlier (Huang, 1999). The FEM solution is built on the multiphysics software FEMLAB® ver. 3.2 (www.comsol.com) in which the membrane vibration is coupled to the acoustic field through "extrusion" coupling variables. The maximum triangular element size is set according to the mean grid size in the ChC solution, L/N or $h/(2M)$, which are made roughly equal in the computation. The type of element used is Lagrange quadratic. The method was successfully used by Chiu *et al.* (2006) for multiphysics coupling of the same silencer involving magnetic forces.

The traditional FGa procedure for the drumlike silencer has been described fully in (Huang, 2002). This procedure is briefly summarized below for the purpose of analyzing its speed of modal convergence. In FGa, the membrane vibration is decomposed into a truncated set of *in vacuo* modes of the simply supported membrane,

$$\frac{\partial\eta}{\partial t}\approx\sum_{j=1}^{N}V_j\sin(j\pi x/L),\quad x\in[0,L], \tag{20}$$

and the modal coefficient of column vector $[V_j]$ are the main unknowns to be determined by the coupling dynamics equation, Eq. (16). The terms of fluid pressure in Eq. (16) have contributions from both the incident wave and that arising from the induced membrane vibration. The latter is calcu-

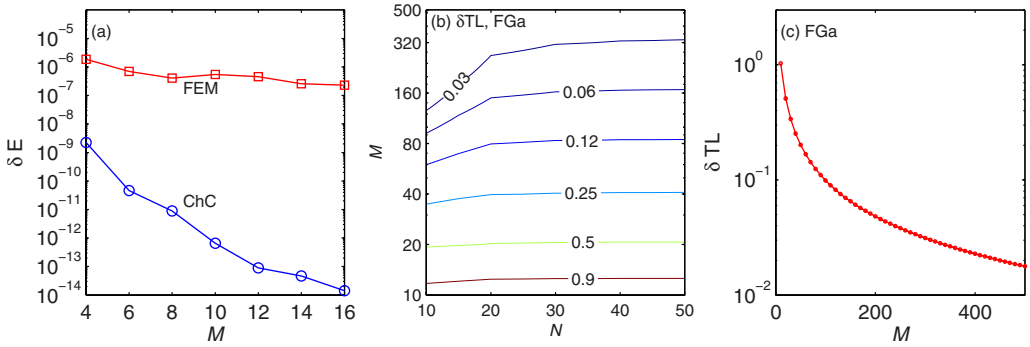


FIG. 3. (Color online) Comparison of convergence for three methods for a drumlike silencer with the parameter setting of Eq. (23) and $f=0.12$. (a) Energy conservation error for FEM and ChC. (b) Contour of the truncation error of FGa measured in TL relative to the most accurate result calculated by ChC with $M=16$. (c) The slow FGa convergence with respect to M for $N=30$.

lated by expansion into duct acoustics modes with rigid walls (Doak, 1973),

$$p(x,y) \approx \frac{\rho_0}{2h}\sum_{n=0}^{M} c_n \psi_n(y) \int_0^L \psi_n(y')V(x') \times [H(x-x')e^{-\iota k_n(x-x')} + H(x'-x)e^{+\iota k_n(x-x')}]dx', \tag{21a}$$

where H is the Heaviside function and

$$\psi_n(y) = \sqrt{2-\delta_{0n}}\cos(n\pi y/h), \quad k_n = -\iota k_0\sqrt{(n\pi/k_0h)^2-1} \tag{21b}$$

are the duct acoustics mode and the modal wavenumber, respectively. The resulting net loading is also expanded into the *in vacuo* vibration modes as in Eq. (20). Here, the loading of vibration mode j contributes to all other modes, $l=1,2,3,\ldots,N$, and the expansion is defined as a modal impedance matrix,

$$\mathcal{Z}_{lj} = \frac{2}{L}\int_0^L (p_{y=0+} - p_{y=0-})_{V=\sin(j\pi x/L)} \sin(l\pi x/L)dx. \tag{22}$$

This matrix allows the vibration induced fluid loading in Eq. (16) to be written as a product of this matrix and the modal amplitude array: $[\mathcal{Z}][V]$ so that Eq. (16) can be solved for the induced vibration amplitudes.

In ChC, the number of axial mesh segments plays much the same role as the number of the membrane vibration modes in FGa, and they are both denoted by N for easy comparison. Likewise, both the number of vertical mesh segments and the duct acoustics modes are denoted by M. When N and M are large, the truncated approximations are expected to converge to the true solution, and the performance of convergence for the three methods is analyzed in Fig. 3. For FGa, N and M are allowed to change in arbitrary combinations. For ChC, N is fixed by M such that the grid sizes in x and y directions are roughly equal.

Numerical tests are first conducted for the following parameters, all given in normalized forms based on the duct height h, fluid density ρ_0, and speed of sound c_0:

$$\frac{h_c}{h}=1, \quad \frac{L}{h}=5, \quad m=\frac{m_0}{\rho_0 h}=1,$$

$$T=\frac{T_0}{\rho_0 c_0^2 h^2}=0.475, \quad \frac{L_U}{h}=\frac{L_D}{h}=2.5. \tag{23}$$

This set corresponds to the optimal design of the drumlike silencer (Huang, 2002).

For the dimensionless frequency of $f=0.12$, tests show that the ChC method converges to a stable value with negligible changes upon further increase in N and M. Figure 3(a) compares the sound energy dissipation for ChC and FEM, which is calculated as follows by integrating the sound intensity fluxes across the inlet and exit, and should vanish in the absence of dissipation mechanisms:

$$\delta E = \int_{\text{inlet}} \frac{1}{2}\text{Re}(pu^*)dy - \int_{\text{exit}} \frac{1}{2}\text{Re}(pu^*)dy,$$

where superscripts * denote complex conjugate. Figure 3(a) shows that the error, δE, decreases rapidly with the mesh number M for ChC while that for FEM drops rather slowly. The best result for FEM is still magnitudes away from that of ChC with the lowest M setting of 4. Figures 3(b) and 3(c) investigate the transmission loss (TL). The value of TL found by ChC for the highest M of 16 is taken as the "exact" solution, denoted as $\text{TL}_{16}=16.374\,43243$ dB, as all other solutions are found to converge to TL_{16} when N and M increase. The deviation from TL_{16} by the FGa solution is shown in Fig. 3(b) as contour plots using N and M as the independent variables. The largely horizontal contour lines imply that the solution depends more strongly on the number of the duct acoustics modes, M, than on the number of the membrane vibration modes, N. In order to achieve an error of δTL smaller than about 0.03 dB, more than 300 duct acoustics modes are required. A cross section of the contour surface is given in Fig. 3(c) for $N=30$, where δTL is shown as a slowly decreasing function of M. The convergence toward high values of M is rather slow. The best FGa solution shown in this figure is roughly comparable to ChC with $M=4$.

In order to illustrate the inadequacy of the FGa method, comparison of solution is made between ChC with $M=16$

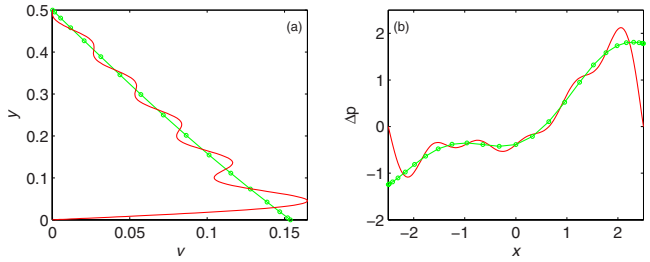


FIG. 4. (Color online) Comparison of FGa (solid line, $M=N=10$) and ChC solutions (-○-, $M=16$) for (a) vertical particle velocity across the duct and for (b) fluid loading along the membrane.

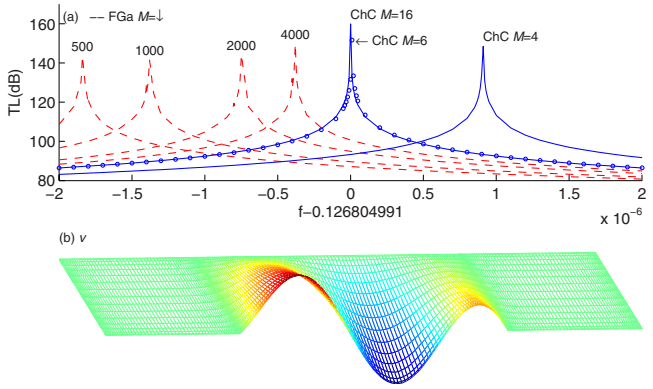


FIG. 5. (Color online) (a) The capturing of a spectral peak using FGa (dashed lines) of various numbers of modes ($M=10N$) and ChC with various M ranging from 4 (solid line) and 6 (open circles) to 16 (solid line). (b) Vertical acoustic particle velocity distribution in the duct for the peak frequency of 0.126 804 991 with TL=159.924 dB.

and FGa with a very low number of modes: $N=M=10$. The vertical velocity v and fluid loading, $\Delta p=p_{y=0-}-p_{y=0+}$, are shown in Figs. 4(a) and 4(b), respectively, with the ChC solution shown in lines with open circles and the FGa solution in a plain solid line. Figure 4(a) shows that the wiggling of v from the FGa solution near the membrane is drastic since the rigid wall duct acoustics mode fails right on the membrane surface, which is nominally located on $y=0$. This failure is caused by the duct acoustics mode expansion of Eqs. (21a) and (21b), where p is expanded in terms of $\cos(n\pi y/h)$. Such expansion implies that v, which can be calculated by the momentum equation $v=-(\rho_0\iota\omega)^{-1}\partial p/\partial y$ and whose value at $y=0$ should be equal to the membrane vibration velocity, is expanded in terms of $\sin(n\pi y/h)$, which, unfortunately, vanishes on $y=0$. Figure 4(b) shows a similar end-point failure for the fluid loading on the membrane expanded in terms of the membrane vibration modes, Eq. (22). This problem may be avoided if the velocity is expanded by functions that do not vanish on the membrane for the silencer region. Different expansions for different regions would eventually lead to a mode-matching scheme, which requires the finding of the eigenfunctions. One such solution for the similar configuration was recently given by Lawrie and Guled (2006).

As is shown in (Huang, 2002), incident sound waves are fully reflected at a few frequencies identified as peaks in the TL spectrum. These spectral peaks do not correspond to any peak in the plot of the kinetic energy of the induced membrane vibration integrated over the membrane length. When the system has no loss mechanism, TL=∞ at these peaks is expected. Figure 5(a) compares the simulation of one such spectral peak by the two methods of various mesh settings. The solid curve with open circles in its proximity is calculated by the ChC method with $M=16$, and is considered to be the most accurate with the precise peak at $f=0.126\,804\,991$ and TL=159.924 dB. The open circles are calculated by ChC with $M=6$. Very close agreement with the result of $M=16$ is reached in terms of the spectral distribution. When M is reduced to 4, the peak position is shifted to $f=0.126\,805\,861$, a deviation of 8.7×10^{-7}, which can be seen clearly on the zoom-in plot of Fig. 5(a). In other words, the convergence of results from $M=4$ to that of $M=16$ takes only a small increase from $M=4$ to $M=6$. Other dashed lines are derived from the FGa method with different numbers of expansion modes and they are explicitly labeled in the figure by their numbers of the duct acoustics modes M, while the number of membrane modes is taken as $N=M/10$ for each calculation. The spectral peak gradually approaches the result of ChC with $M=16$. The number of duct acoustics mode, M, needed to reach an accuracy level of frequency deviation of around 10^{-6} is several thousands. This is another indication of how good the convergence is for the ChC method. When complete sound reflection occurs, a question arises as to whether some form of rapid variation exists in the acoustic field such that it becomes difficult for the current spectral methods to simulate. The results of the peak calculated by ChC with $M=16$ are shown in Fig. 5(b) in terms of the vertical acoustic particle velocity v. The field is found to be completely smooth. The physical reason for such smoothness is found in the near field, which spreads out from the membrane region.

The performance of the ChC is now further compared with the FEM for a dissipative system at a much higher frequency. The configuration used here is similar to the drum silencer studied above but with several modifications. As shown in Fig. 1, the cavity is now filled with sound absorption material, which is described in the next section as equivalent fluid with complex density and speed of sound. The upstream boundary at $x=-L_U$ is proscribed with a distribution of axial vibration speed of $\phi_x|_{x=-L_U}=f(y)$ that is compatible with the solid wall conditions on $y=0$, $h/2$. The exit boundary at $x=L+L_D$ is set as a rigid wall. The frequency used for the following example is $f=3.6765$, which is derived from a rounded dimensional frequency of 2500 Hz in a duct of 0.5 m in height. The acoustic field has propagating modes up to the seventh order in both upstream and downstream regions for the full-height duct or to the third-order mode in terms of the half-height duct. Despite such high frequency, there is no uncertainty in the proscribed boundary condition in this configuration, and both FEM and ChC are found to converge to the same solution. All parameters used for the configuration are listed as follows:

$$L_U=L_D=2.5h,\quad L=5h,\quad h_c=h/2,\quad \phi_x|_{\text{inlet}}=\cos(4\pi y/h),\quad m=0.1,\quad T=0.35,\quad R=0.5, \tag{24}$$

where dimensionless flow resistivity R and other sound ab-

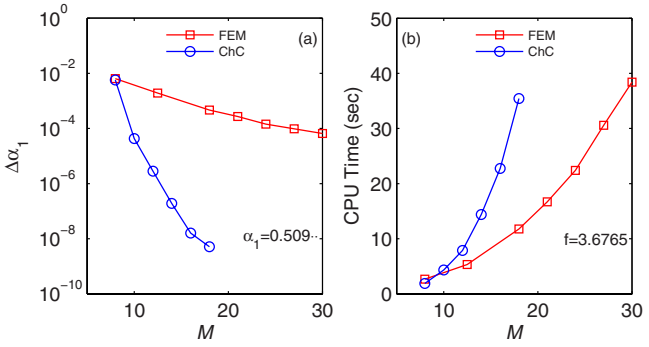


FIG. 6. (Color online) Comparison of (a) accuracy and (b) CPU time performance between ChC and FEM for a dissipative system at high frequency. The parameters are set in Eq. (24).

sorption material properties are defined in Eq. (25) below. The ratio of sound energy absorption, α_1, is defined as the sound energy flux through the membrane divided by the power that would have been radiated by the specified inlet vibration in a uniform duct of semi-infinite length. The ChC mesh is characterized by the number of the cross sectional Gauss–Lobatto points, M, while the maximum side length of the quadratic elements used in FEM is limited to $1/M$ of the half of duct height. As M increases, α_1 predicted by ChC converges to 0.509 1147 561 84 15, and the deviation from this value, denoted as $\Delta\alpha_1$, is around 5×10^{-9} when M is 18, and is around 1×10^{-4} for FEM when $M=30$. The same occurs for the difference between two solutions with neighboring even integers of M. Figure 6(a) shows the decrease of $\Delta\alpha_1$ with M, and Fig. 6(b) compares the CPU time spent on a Core 2 Duo® processor with a 2.5 GHz clock speed. The absolute minimum number of grid to resolve a wavelength is π for the Gauss–Lobatto mesh, which means $2M \geqslant \pi f$, $M \geqslant \pi f/2=5.77\approx6$. Figure 6 shows that the performance of the ChC is almost identical to that of the FEM for $M=8$ for both accuracy and CPU time (~2 s). However, when $M=10$ is used for the ChC, its accuracy is already better than that of the FEM at $M=30$, while its CPU time of 4.3 s is around 11% of the FEM at $M=30$, or the same CPU time as the FEM at around $M=11$.

III. ABSORPTION OF HIGH-FREQUENCY SOUND

The aspects of prediction methodology specific to the absorption of high-frequency sound are given in Sec. III A. Optimization for both pure duct lining and membrane covered lining is described in Sec. III B. Sec. III C is devoted to the analysis of why the membrane cover helps the absorption of high-frequency sound in grazing incidence. In this section, numerical examples are given for the cavity depth of $h_c=h/2$, shallower than that in Sec. II.

A. Modeling of porous material and outgoing waves

The sound absorption material in the cavity is modeled by an equivalent homogeneous fluid with complex density and speed of sound. The empirical regression formulas given by Delany and Bazley (1970) are used here for simulation purposes. They are listed as follows together with definitions of normalized frequency f_{sam} in the sound absorption material and flow resistivity R:

$$\frac{\rho c}{\rho_0 c_0} = 1 + 0.0571 f_{\text{sam}}^{-0.754} - \iota \times 0.087 f_{\text{sam}}^{-0.732},$$

$$\frac{c_0}{c} = 1 + 0.0978 f_{\text{sam}}^{-0.700} - \iota \times 0.189 f_{\text{sam}}^{-0.595},$$

$$R = \frac{R_0 h_c}{\rho_0 c_0}, \quad f_{\text{sam}} = \frac{\rho_0 f^*}{R_0} = \frac{f}{R} \times \frac{h_c}{h}, \tag{25}$$

where R_0 is the dimensional flow resistivity. As pointed out by Delany and Bazley (1970), these empirical formulas can be used with confidence for $f_{\text{sam}} \in [0.012, 1.25]$. Although they also had some experimental data beyond this range, problems occur for very low frequencies where unphysical properties, such as negative real part of complex density, $\text{Re}(\rho)<0$. It can be shown that $\text{Re}(\rho)=0$ occurs at $f_{\text{sam}}=0.0025$ based on Eq. (25). Remedies are available, e.g., (Kirby and Cummings, 1999), but this problem is not seriously encountered in the following calculations. The optimal R for the traditional duct lining is found to be $R_{\text{opt}}=40.1$. For the lowest dimensionless frequency of $f=0.3$, this gives a minimum $f_{\text{sam}}=0.0037$ and the corresponding real part of density is $\text{Re}(\rho/\rho_0)=1.30$. The corresponding values for the optimization of the membrane covered lining are $R_{\text{opt}}=0.39$, $\min(f_{\text{sam}})=0.038$, $\text{Re}(\rho/\rho_0)=1.27$.

When the frequency of the incident wave exceeds the first cut-on frequency of the duct, $c_0/2\,h$, or $f=0.5$, higher-order modes propagate. The plane-wave boundary conditions, Eqs. (3a) and (3b), no longer apply at the inlet and exit, respectively. Within a duct, the wave operator may be separated into left- and right traveling components,

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{c^2 \partial t^2}\right) p = 0 \rightarrow \underbrace{\left(\frac{\partial}{\partial x} + \sqrt{\frac{\partial^2}{c^2 \partial t^2} - \frac{\partial^2}{\partial y^2}}\right)}_{\text{waves to } +x}$$
$$\times \underbrace{\left(\frac{\partial}{\partial x} - \sqrt{\frac{\partial^2}{c^2 \partial t^2} - \frac{\partial^2}{\partial y^2}}\right)}_{\text{waves to } -x} p = 0. \tag{26}$$

The square-root terms represent pseudodifferentiations that can only be approximated by, for instance, rational functions of polynomials in a Padé approximation: $P_m(x)/P_n(x)$, where $P_m(x)$ and $P_n(x)$ are polynomial functions of orders m and n, respectively, and the corresponding scheme is called a (m,n) scheme. Different orders of accuracy are obtained with different polynomials, and each approximation may allow near-perfect pass of outgoing waves through an artificial (computational) boundary in a certain angular range with respect to the normal direction to the boundary (Halpern and Trefethen, 1988). For example, the simplistic Taylor expansion of the square-root function $\sqrt{1-x^2}\approx 1-\frac{1}{2}x^2$ can be used to approximate the dispersion relation from Eq. (26), in a (2,0) Padé scheme. This scheme allows waves to pass through a bound-

ary within 15° of angular range, and the corresponding one-way wave equation (for $+x$) is

$$\text{Padé}(2,0)\text{scheme:}\frac{\partial^2 p}{c_0 \partial x \partial t} + \frac{\partial^2 p}{c_0^2 \partial t^2} - \frac{1}{2} \times \frac{\partial^2 p}{\partial y^2} = 0, \quad (27)$$

which will be used as a low-order scheme for comparison purposes. Many numerical schemes exist to implement such nonreflecting boundary conditions, and they may be categorized into local and nonlocal schemes (Givoli, 2001) in terms of whether the evaluation of the condition is coupled to the values on the whole boundary. The so-called DtN condition (Keller and Givoli, 1989) relates p with its normal derivative, p_x, on the boundary and is an exact but nonlocal condition. In the current context, that relationship is essentially an acoustic impedance in the near field for waves propagating in the angular range of $(-\pi/2, \pi/2)$ as well as for evanescent waves. A nonlocal condition like this can spoil the bandedness of the stiffness matrix in FEM and thus many local outgoing wave conditions of high order of accuracy have been developed (Givoli, 2004), the perfectly matched layer (PML) being another alternative with similar advantages (Singer and Turkel, 2004).

However, for the current problem of sound-membrane interaction, strong coupling is already present, and the use of a nonlocal DtN condition imposes no extra coding or computational burden of strong coupling between boundary and interior points. In its discrete form, derivatives inside the square-root of Eq. (26) can be represented by the Chebyshev derivative matrices, for which the first order is given in Eq. (6b). The square-root of a matrix is calculated by eigenvalue decomposition given in Eq. (10),

$$\left[\frac{\partial^2}{c^2 \partial t^2} - \frac{\partial^2}{\partial y^2}\right] = k_0^2[I] - [D^{(yy)}] = \mathcal{K}\Lambda\mathcal{K}^{-1},$$

$$K_{\text{sq}} \equiv \left[\sqrt{\frac{\partial^2}{c^2 \partial t^2} - \frac{\partial^2}{\partial y^2}}\right] = \mathcal{K}\Lambda^{1/2}\mathcal{K}^{-1}. \quad (28)$$

When the square-root of nonpositive eigenvalue arises, the choice is made to pick the decaying wavenumber. The inlet boundary condition is simply the left-traveling wave equation for $(p - p_{\text{in}})$, where p_{in} is the specified incident wave given in Eq. (2), hence,

$$[d^{(x)}][p - p_{\text{in}}] - K_{\text{sq}}[p - p_{\text{in}}] = 0, \quad (29)$$

while the exit boundary condition is

$$[d^{(x)}][p] + K_{\text{sq}}[p] = 0. \quad (30)$$

The values of p at the inlet and exit boundaries are unknowns to be found together with values on other interfaces, namely, ξ, χ, and μ in Eqs. (18a) and (18b), by line forcing perturbations on the inlet and exit boundaries as described in the paragraph above Eqs. (14a) and (14b), and incorporate Eqs. (29) and (30) into Eqs. (18a) and (18b).

The performance of the above DtN implementation is studied below by comparing with the analytical solution. Consider a square domain of dimension $\pi \times \pi$, with pressure release condition, $p=0$, at the bottom $(y=0)$ and top boundaries $(y=\pi)$. A wave of wavenumber k and cross sectional mode n is launched at the left-hand side, $x=0$, and the task is to propagate this wave out of the right-hand side of the domain, $x=\pi$, without reflection. The analytical solution for the problem is simply

$$p = \sin(ny)\exp(-\imath x\sqrt{k^2 - n^2}). \quad (31)$$

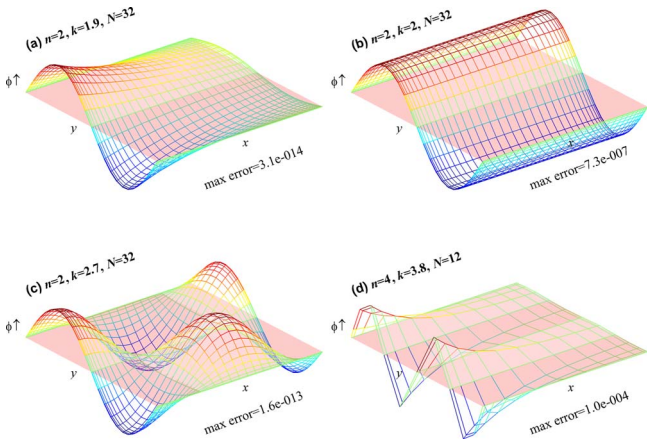


FIG. 7. (Color online) Calibration of the DtN boundary condition by comparison with the known solution, Eq. (30), for the higher-order duct acoustics modes. The square domain of size $\pi \times \pi$ is discretized by $N \times N$ Gauss–Lobatto grids.

The computed results are shown in Fig. 7 with deviation from the above exact solution given in the label along the x axis in each subfigure. Figure 7(a) shows an evanescent wave of second mode with $k=1.9$ and $N=32$ Gauss–Lobatto grids. The maximum error is 3.1×10^{-14}. When $n=k$, the wave is almost parallel with the exit boundary. Local schemes of outgoing wave would normally fail to propagate the wave out of the domain (Halpern and Trefethen, 1998), but the current DtN scheme with $N=32$ gives a solution with a maximum error of 7.3×10^{-7}, as shown in Fig. 7(b). Figure 7(c) shows the propagating second mode at $k=2.7$, and the maximum error is also very low: 1.6×10^{-13}. Figure 7(d) explores the performance of a very coarse mesh, 12 grids for the fourth-order evanescent wave, and the maximum error is 1.0×10^{-4}. As a reference, comparison is also made with a scheme of PML by Singer and Turkel (2004) with $n=6$, $k=8$. The current scheme with 32 grids gives a maximum error of 1.02×10^{-13} while the most accurate PML scheme in that study with 128 grids gives 4.71×10^{-7}. As another reference, the Padé (2,0) scheme of Eq. (27) gives a maximum error of 8.20×10^{-2} with $N=32$ Gauss–Lobatto grids and Chebyshev derivatives.

B. Optimization of duct lining

Focusing on the absorption of high-frequency noise, the incident wave is assumed to have a uniform SPL from dimensionless frequency of $f=0.3-3.0$. Note that $f=0.5$ signifies the first cut-on frequency of the rigid walled section of the duct, and $f=3$ means the cut-on of the sixth-order mode. For the purpose of our optimization study, there is no noise outside this frequency range. A-weighting is used to calculate the overall noise attenuation (TL_Σ). The normalized attenuation rate is defined as

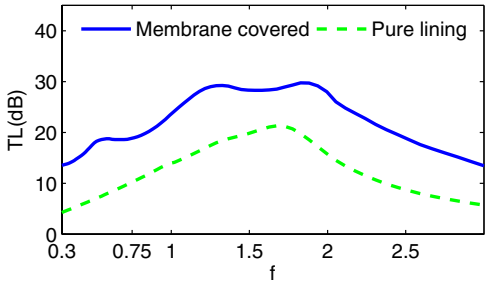


FIG. 8. (Color online) Spectral comparison between the optimal membrane covered duct lining with $m=0.1$, $T=0.361$, $R=0.39$ and the optimal pure duct lining with $R=40.1$.

$$L_h = TL_\Sigma/(2L/h) \tag{32}$$

The incident wave is a plane wave, and the strength of the transmitted wave is calculated in terms of the time-mean sound power flux integrated over the duct cross section. For A-weighting to be applied, the dimensionless frequency has to be converted to the dimensional one. In this exercise, a duct of height $h=0.5$ m is chosen, so that 1 kHz corresponds to dimensionless frequency $f=f^*h/c_0=10^3 \times 0.5/340=1.4706$. For a pure duct lining, the dimensionless flow resistivity R defined in Eq. (25) is the only controlling parameter other than the cavity geometry. The result of optimization for a cavity geometry of $L=5\ h$ and $h_c=0.5\ h$ is that $TL_\Sigma=10.6$ dBA is achieved with $R=40.1$.

When a membrane cover is added, the tensile force on the membrane, T_0, becomes the second controlling parameter when the membrane-to-air mass ratio is fixed. For the purpose of investigating the effect of dimensionless tension $T=T_0/(\rho_0 c_0^2 h^2)$, a very light membrane with mass ratio of $m=m_0/(\rho_0 h)=0.1$ is used. The optimal combination is found to be $T=0.361$ and $R=0.39$. The optimal overall noise attenuation is $TL_\Sigma=19.8$ dBA, which is 9.2 dBA higher than the pure duct lining. Figure 8 compares the optimal membrane covered lining (solid line) with that of the pure lining (dashed line). The spectral improvement is almost constant for the whole frequency range.

In terms of L_h, the improvement from the pure duct lining to the best membrane lining is $1.98-1.06=0.92$ dBA. If these attenuation rates are regarded as pure absorption, the overall absorption rate is calculated by $(1-10^{-L_h/10})$ per $h/2$. This works out to be 21.6% and 36.6%, which represent an increase of $(36.6/21.6-1)100\%=69\%$.

C. Mechanism of the membrane cover effect

The exact mechanism by which the membrane cover enhances the absorption of high-frequency sound is analyzed for the best setting for $m=0.1$, $L=5h$. Figure 9 shows the SPL and $|v|$ distributions for the frequency of $f=0.57$. Figure 9(a) shows the SPL distribution in the main duct. The upstream region is rather flat, implying the absence of a strong standing wave or the lack of significant wave reflection. The downstream region is also rather flat, meaning the higher-order mode is not strong in the exit region. Figure 9(b) shows the vertical component of the particle velocity v for both duct and cavity.

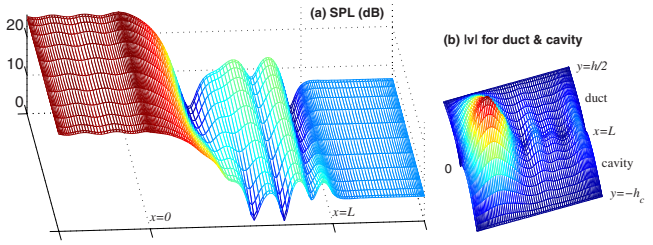


FIG. 9. (Color online) Acoustic field for a membrane covered duct lining for the best setting for $m=0.1$, $L/h=5$, and $f=0.57$. (a) SPL distribution in the duct passage. (b) Vertical velocity distribution in the duct (upper) and cavity (lower) regions, joined together by the membrane interface.

Figure 10 shows the membrane cover effect by comparing the sound intensity and cross sectional mode distribution for pure lining (dashed lines) and membrane covered lining (solid lines, same parameter setting as Fig. 9). Figure 10(a) shows that the membrane vibration near the leading edge essentially drains most incident noise into the cavity in a rather concentrated manner, while pure duct lining does so gradually with much less effectiveness. Figure 10(b) shows the decay of the plane-wave mode with a ripple in the upstream region due to a weak standing wave. Figure 10(c) is for the first-order mode ($n=1$) for the half-width passage (or the second-order mode for the full passage) defined as

$$A_n(x) = \frac{4}{h}\int_0^{h/2} p(x,y)\cos(2ny\pi/h)dy. \tag{33}$$

For the membrane covered lining, the sudden decrease of the plane-wave mode near the leading edge of the membrane is accompanied by the sudden increase of the first mode coefficient A_1. In other words, the membrane scatters the plane wave into higher-order modes. This is the essence of the membrane cover enhancement of the high-frequency sound absorption.

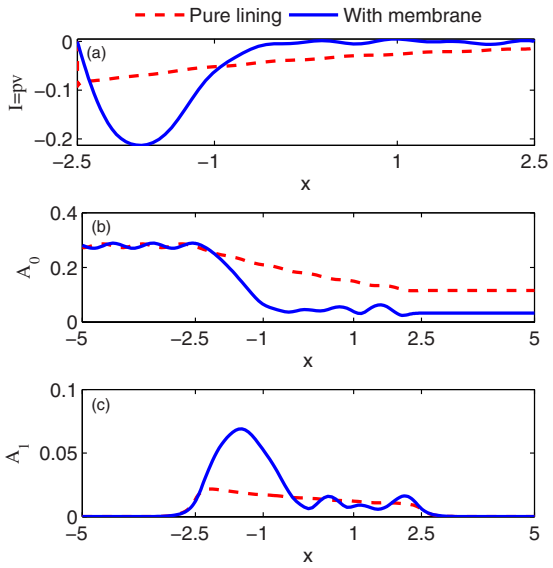


FIG. 10. (Color online) Analysis of the effect of membrane on the lining. (a) Sound intensity across the membrane interface. (b) Plane-wave mode ($n=0$) variation with distance. (c) First mode ($n=1$) for the half-passage as defined by Eq. (33).

IV. CONCLUSIONS

The ChC method converges very rapidly and shows great accuracy when compared with a FD4. The characteristic exponential convergence to the exact solution is achieved with a very moderate mesh density. Favorable comparison is also achieved with finite element modeling.

The capability of the ChC to handle a nonperiodic boundary condition contrasts with the Fourier expansion utilizing the rigid walled duct acoustics modes. In the latter method, the modal expansion fails on the vibrating boundaries leading to a contrived and very slow convergence.

The method of domain decomposition can be conveniently adopted for this linear problem to limit the mesh size of each domain. The size of the "stiffness" matrix involved is the total number of mesh points on all interfaces, which have to be resolved simultaneously. The DtN outgoing wave condition is naturally embedded in such domain decomposition scheme, and the nonlocalness of the condition does not impose any extra coding or computational burden as it might do to standard FEM.

The method is applied to predict the enhancement of the absorption of high-frequency sound by a membrane cover on the traditional lining. When a suitable tension is applied on the cover, the plane incident wave is scattered into higher-order modes, which then enter into the lining more easily and get dissipated. The scattering mainly takes place near the leading edge of the membrane.

When optimized for the noise absorption from dimensionless frequency $f=0.3-3$, 0.5 being the first cut-on frequency of the central duct, the membrane cover is found to offer an additional 0.9 dB attenuation per axial distance equal to half of the duct passage height.

ACKNOWLEDGMENTS

The author thanks the Research Grants Council of the Hong Kong SAR Government for its funding of two related research projects (HKU 5298/03E and HKU5170/05E).

Bartels, R. H., and Stewart, G. W. (**1972**). "Algorithm 432: Solution of the matrix equation AX+XB=C," Commun. ACM **15**, 820–826.

Bayliss, A., Class, A., and Matkowsky, B. J. (**1994**). "Roundoff error in computing derivatives using the Chebyshev differentiation matrix," J. Comput. Phys. **116**, 380–383.

Boyd, J. P. (**2001**). *Chebyshev and Fourier Spectral Methods*, 2nd ed., Dover, New York.

Cheng, C. Y. R., Seybert, A. F., and Wu, T. W. (**1991**). "A multidomain boundary element solution for silencer and muffler performance prediction," J. Sound Vib. **151**, 119–129.

Chiu, Y. H., Cheng, L., and Huang, L. (**2006**). "Drum-like silencers using magnetic forces in a pressurized cavity," J. Sound Vib. **297**, 895–915.

Choy, Y. S., and Huang, L. (**2002**). "Experimental studies of a drum-like silencer," J. Acoust. Soc. Am. **112**, 2026–2035.

Choy, Y. S., Huang, J., Huang, L., and Zhou, Y. (**2007**). "An experimental study of flow induced vibration on a tensioned membrane," J. Mech. Sci. Tech. **21**, 1359–1366.

Clenshaw, C. W., and Curtis, A. R. (**1960**). "A method for numerical integration on an automatic computer," Numer. Math. **2**, 197–205.

Cummings, A. (**1999**). "High frequency ray acoustics models for duct silencers," J. Sound Vib. **221**, 681–708.

Delany, M. E., and Bazley, E. N. (**1970**). "Acoustical properties of fibrous materials," Appl. Acoust. **3**, 105–116.

Doak, P. E. (**1973**). "Excitation, transmission and radiation of sound from source distributions in hard-walled ducts of finite length. I. The effects of duct cross-section geometry and source distribution space-time pattern," J. Sound Vib. **31**, 1–72.

Givoli, D. (**2001**). "High-order nonreflecting boundary conditions without high-order derivatives," J. Comput. Phys. **170**, 849–870.

Givoli, D. (**2004**). "High-order local non-reflecting boundary conditions: A review," Wave Motion **39**, 319–326.

Gottlieb, D., Hussaini, M. Y., and Orszag, S. A. (**1984**). "Introduction: Theory and applications of spectral methods," in *Spectral Methods for Partial Differential Equations*, edited by R. G. Voigt, D. Gottlieb, and M. Y. Hussaini, SIAM, Philadelphia, PA.

Halpern, L., and Trefethen, L. N. (**1988**). "Wide-angle one-way wave-equations," J. Acoust. Soc. Am. **84**, 1397–1404.

Horoshenkov, K. V., and Sakagami, K. (**2001**). "A method to calculate the acoustic response of a thin, baffled, simply supported poroelastic plate," J. Acoust. Soc. Am. **110**, 904–917.

Huang, L. (**1999**). "A theory of passive duct noise control by flexible panels," J. Acoust. Soc. Am. **106**, 1801–1809.

Huang, L. (**2002**). "Modal analysis of a drum-like silencer," J. Acoust. Soc. Am. **112**, 2014–2025.

Ingard, K. U. (**1994**) *Notes on the Sound Absorption Technology*, Noise Control Foundation, Poughkeepsie, New York.

Karniadakis, G. E., and Sherwin, S. (**2005**). *Spectral/hp Element Methods for Computational Fluid Dynamics*, 2nd ed., Oxford University Press, Oxford.

Keller, J. B., and Givoli, D. (**1989**). "Exact non-reflecting boundary conditions," J. Comput. Phys. **82**, 172–192.

Kirby, R., and Cummings, A. (**1999**). "Prediction of the bulk acoustic properties of fibrous materials at low frequencies," Appl. Acoust. **56**, 101–125.

Kirby, R., and Lawrie, J. B. (**2005**). "A point collocation approach to modeling large dissipative silencers," J. Sound Vib. **286**, 313–339.

Kosloff, R., and Tal-Ezer, H. (**1993**). "Modified Chebyshev pseudospectral methods with $O(N^{-1})$ time step restriction," J. Comput. Phys. **104**, 457–469.

Lawrie, J. B., and Guled, I. M. M. (**2006**). "On tuning a reactive silencer by varying the position of an internal membrane," J. Acoust. Soc. Am. **120**, 780–790.

Lawrie, J. B., and Kirby, R. (**2006**). "Mode-matching without root-finding: Application to a dissipative silencer," J. Acoust. Soc. Am. **119**, 2050–2061.

Lynch, R. E., Rice, J. R., and Thomas, D. H. (**1964**). "Direct solution to partial differential equations by tensor product method," Numer. Math. **6**, 185–199.

Mechel, F. P. (**1990**). "Numerical results to the theory of baffle-type silencers," Acustica **72**, 7–20.

Munjal, M. L., Galaitsis, A. G., and Vér, I. L. (**2006**) "Passive silencers," in *Noise and Vibration Control Engineering: Principles and Applications*, edited by L. L. Beranek and I. L. Vér, Wiley, Hoboken, NJ, Chap. 9.

Munjal, M. L., and Thawani, P. T. (**1997**). "Effect of protective layer on the performance of absorptive ducts," Noise Control Eng. J. **45**, 14–18.

Patera, A. T. (**1984**). "A spectral element method for fluid dynamics: laminar flow in a channel expansion," J. Comput. Phys. **54**, 468–488.

Peat, K. S., and Rathi, K. L. (**1995**) "A finite-element analysis of the convected acoustic-wave motion in dissipative silencers," J. Sound Vib. **184**, 529–545

Peyret, R. (**2002**). *Spectral Methods for Incompressible Viscous Flow*, Springer-Verlag, New York.

Singer, I., and Turkel, E. (**1998**). "High-order finite difference methods for the Helmholtz equation," Comput. Methods Appl. Mech. Eng. **163**, 343–358.

Singer, I., and Turkel, E. (**2004**). "A perfectly matched layer for the Helmholtz equation in a semi-infinite strip," J. Comput. Phys. **201**, 439–465.

Annoyance responses to stable and changing aircraft noise exposure

Mark Brink,[a] Katja E. Wirth,[b] and Christoph Schierz[c]
ETH Zurich, MTEC-ZOA Public and Organizational Health, CH-8092 Zurich, Switzerland

Georg Thomann[d]
Empa Materials Science & Technology, Acoustics Laboratory, CH-8600 Dübendorf, Switzerland

Georg Bauer[e]
University of Zurich, Institute of Social and Preventive Medicine, CH-8001 Zurich, Switzerland

(Received 21 March 2008; revised 16 June 2008; accepted 29 July 2008)

This article reports the two extensive aircraft noise annoyance surveys subsequently carried out among residents in the vicinity of Zurich Airport in 2001 and 2003 in order to update and validate existing exposure-effect relationships for aircraft noise and annoyance in Switzerland. Logistic and polynomial approximations of the exposure-annoyance relationships for both the years 2001 and 2003 are presented for the L_{dn}, L_{den}, and $L_{A,eq24}$ noise metrics. The results confirm other recently published international research and provide further evidence that community annoyance due to aircraft noise has increased over the past decades. Between the two survey years, a considerable amount of early morning and late evening flight operations have been relocated to use an other runway than before; thus both the effects of a recent step decrease and recent step increase on the exposure-annoyance relationship could be investigated. Residents that experienced a step increase elicited a quite pronounced over-reaction of annoyance which correlated with the magnitude of the change. Two logistic regression models are provided to forecast the effects of changes in exposure during shoulder hours in the early morning and the late evening.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977680]

PACS number(s): 43.50.Qp, 43.50.Lj [BSF] Pages: 2930–2941

I. INTRODUCTION

In areas with day-night aircraft noise levels (L_{dn}) below 70 dB(A), annoyance is among the most important health effects caused by aircraft noise (Berglund and Lindvall, 1995). A wealth of exposure-effect functions for noise annoyance has been established in the past decades, relating traffic noise exposure, including aircraft noise, to the percentage of highly annoyed persons (Fidell *et al.*, 1991; Miedema and Vos, 1998; Schultz, 1978). Some of them are based on very large data sets, collected from different studies in various countries (Miedema and Vos, 1998), and for many years defined a sort of *de facto* standard for noise impact assessment for noise policy issues, e.g., within the scope of noise abatement in the European Union (European Commission, 2002). However, many of the original data sets with studies dating back as far as to the 1960s are by now to be considered outdated and may not any longer correctly reflect the relationship between noise metrics and annoyance measures. There is evidence that today, people are more sensitive toward aircraft noise than they were decades ago. Guski (2004) reanalyzed the data from the Miedema and Vos meta-analysis (1998) with respect to the year of the study and the respective L_{dn} for 25% highly annoyed (HA) persons and found that the exposure level needed to elicit a particular level of annoyance decreased considerably over the past decades. This trend has also been investigated and confirmed by van Kempen and van Kamp (2005) who added more recent studies to the data set. As could impressively be demonstrated in a recent multinational study (Babisch *et al.*, 2007), the annoyance shift seems to be specific to aircraft noise: Whereas the so called "EU curve" (European Commission, 2002) for road traffic noise very well matches the exposure-effect relationships that can be found with new survey data, the EU curve for aircraft noise systematically underestimates the percentage of HA persons at any given exposure level. Concerning the reasons of this shift of the exposure-effect curve, several explanations are being currently discussed. On the one hand, in the past two decades, as Bröer and Wirth (Bröer, 2007; Bröer and Wirth, 2004) argued on the background of Beck's (1992) theory of risk society, aviation is no longer considered a sign of modernity and technical progress, and probably steadily lost its technological advancement appeal to its adverse effects such as noise and air pollution. On the other hand, the numbers of air traffic movements have doubled or tripled at many airports in the past decades, whereas the sound energy of single aircraft movements has consistently decreased, thus altering the trade-off between number of movements and total sound energy of all movements which might also lead to a change in the overall perception of aircraft noise and ultimately to an increase of annoyance. It has been reported as a result of the

[a]Electronic mail: brink@ethz.ch
[b]Electronic mail: katja.wirth@euresearch.ch
[c]Electronic mail: cschierz@ethz.ch
[d]Electronic mail: georg.thomann@empa.ch
[e]Electronic mail: gfbauer@ ifspm.uzh.ch

recently published ANASE study in the UK that the number of aircraft movements today better explains variance in annoyance than it did 20 years ago (Department for Transport, 2007). Further possible explanations for this trend, e.g., increasing public debate about the continuous expansion plans of airports, have been discussed in van Kempen and van Kamp (2005). In conclusion, in noise effect research, routinely updating the databases for establishing exposure-effect functions that provide a sound basis for noise impact assessment for current and future scenarios is a permanent necessity. When it comes to forecasting community response to aircraft noise after a (prospective) operational change (e.g., opening of a new runway), an additional source of uncertainty in the estimation of the degree of annoyance potentially comes into play, the so called over-reaction effect. The current study aims to tackle both these issues.

A. Influence of a step change in exposure

A typical application for exposure-effect functions that are derived from community surveys would be when public authorities must forecast future annoyance in the course of a prospective airport expansion, opening of a new airport, or changes of the flight regime or operating plan. As such parameters change, they usually elicit a step change in exposure. There is evidence that a step change of noise exposure generally goes along with a so called "over-reaction" of the residents: With an increase of the exposure level, people are more annoyed than would be predicted by steady-state exposure-effect curves, whereas with a decrease of the level, they are less annoyed than would be predicted by the same curves (Fidell *et al.*, 1985; Fidell *et al.*, 2002; Griffiths and Raw, 1989). Although some models and tentative explanations have previously been published (Job, 1988b; Job and Hatfield, 2003), the mechanism of how residents judge their level of annoyance in response to the exposure change is not understood in detail. As pointed out by Fidell *et al.* (2002), the lack of information may partly be due to scarce opportunities to carry out field studies on the effect of abrupt and clear changes in noise exposure. It is useful, however, to seek a better understanding of these effects, in order to serve the interest of local governments, airport authorities, and the public in as precise as possible predictions of the effects of (future) changes of noise exposure.

In the current study, we addressed this issue by carrying out two annoyance surveys among residents around Zurich Airport, in Switzerland. The first survey was completed in August 2001, the second—employing almost identical methodology—was carried out two years later, during August 2003. Between these two surveys, in some communities to the east of the airport, the exposure situation changed considerably during shoulder hours (late evening and early morning) after new regulations, reducing air traffic over south German territory, were put into force in October 2001. As a consequence of these new regulations, residents to the east of the airport with no or only little experience in aircraft noise up to that point were newly overflown by inbound aircraft during shoulder hours. This quasiexperimental setting allowed for a detailed analysis of the effects of a step change in exposure to annoyance. To get an overview of the exposure around Zurich Airport and the local exposure change during the study period, Fig. 3 provides a map of the Zurich Airport region with exposure level contours for the year 2003, exposure change as grayscale gradient, and a schematic drawing of the runway layout.

B. Background and study objectives

Rapid and substantial changes in the nature of noise associated with commercial aviation have been a persistent feature of the industry since the 1950s. At Zurich, the total number of air traffic movements doubled between 1980 and 2000, and reached its interim peak in the year 2000 with 325 622 movements. In the aftermath of the demise of the airports most important customer, Swissair, which declared insolvency in October 2001, the number of aircraft movements decreased during the study period (2001: 309 230; 2002: 282 254; 2003: 269 392 movements) but is taking up again nowadays. As has been discussed already, in such a changing context, we cannot assume that people react in the same way to aircraft noise as they did decades ago. The last aircraft noise annoyance surveys carried out in Switzerland date back to 1971 (Arbeitsgemeinschaft Für Sozio-Psychologische Fluglärmuntersuchungen, 1974) and 1991 (Oliva, 1998), respectively. In order to update our knowledge about resident's annoyance and their attitudes toward air traffic, a new series of surveys was carried out in 2001 and 2003.

The main research objectives were the following:

(1) provide an updated exposure-effect function for high annoyance among residents in the vicinity of Zurich Airport;
(2) investigate the impact of a step change (increase/decrease) of aircraft noise exposure which, in some areas, took place between 2001 and 2003; provide a model that estimates the impact of change in exposure on high annoyance.

II. METHODS

A. Questionnaire and telephone interviews

The two surveys were carried out by mailed questionnaires (2001 and 2003) and in parts by computer assisted telephone interviews (2003) commissioned to a market research bureau. Both surveys 2001 and 2003 employed a noise annoyance questionnaire in written form that was mailed each survey year to 3360 addresses in the vicinity of the airport. Answers were collected between mid-August and September 10 (2001) and mid-October (2003), respectively. The respondents' addresses were selected according to a random procedure (described in Sec. II C). Respondents were asked to rate, among other aspects, their aircraft noise annoyance. For this purpose, the questionnaire included the German version of the 11-point annoyance scale from 0 to 10 recommended by ICBEN (Fields *et al.*, 2001), as well as 7-point verbal scales with marks from "not annoyed at all" to "extremely annoyed," for rating different situations. The 11-point annoyance questions referred to the civil aircraft noise experience of the respondents during the 12 preceding

months. For both the 2001 and the 2003 surveys, the same questionnaire was used. A total of 510 telephone interviews were carried out in 2003, using the same order of questions and identical wording as in the written questionnaire form. Within the scope of a mixed model analysis described in Sec. III B, we found no evidence for an effect of survey method (questionnaire versus telephone interviews) on annoyance. We could therefore conclude that the cost-saving method of mail-distributed questionnaires is adequate for a study with the current aim and scope.

B. Acoustical noise indicators

For each respondent in both surveys, extensive noise exposure calculations based on the exact geographical location of the respondent's location of residence were made. The geographical coordinates for the modeling of the noise exposure were derived from the respondent's address using a geographical information system (GIS). The calculations were performed using the aircraft noise simulation model FLULA2 developed at Empa Materials Science and Technology (Krebs *et al.*, 2004; Pietrzko and Hofmann, 1988; Thomann, 2007). The model delivers aircraft noise exposure data for an untilled terrain on a regular grid of 250 m. For each respondent's place of residence, the exposure data were linearly interpolated. Noise exposure assessment with FLULA2 delivers the distribution of maximum sound pressure levels, the average sound pressure level L_{Aeq} for different time periods, as well as L_{den} and L_{dn}. In order to be able to account for 1 h-L_{Aeq} values and their influence on time-of-day-dependent noise sensitivity (preliminary results in Brink *et al.*, 2007), for all respondents the 24 yearly average 1 h-L_{Aeq} values were calculated according to

$$L_{Aeq,Hi} = 10\log\left(\frac{t_0}{t}\sum_k\sum_j N_{Hjk}\times 10^{0.1\cdot L_{AEijk}}\right) = 10\log\left(\sum_k\sum_j N_{Hjk}\times 10^{0.1\cdot L_{AEijk}}\right) + 10\log\left(\frac{t_0}{t}\right), \quad (1)$$

where L_{AE} is the calculated, averaged single exposure level in decibel at a specific immission point i, $L_{Aeq,Hi}$ is the A-weighted average sound pressure level for the hour of the day H at immission point i, t is the averaging time in seconds, for any hour 3600 s, t_0 is the reference time in seconds, H is the hour of the day (example: $H=0$ is the hour from 00:00 to 00:50:59; $H=19$ is hour 19:00 to 19:59:59), N is the number of movements during the averaging time, j is the index for a specific aircraft type, k is the index for a specific flight route, and i is the index for a specific immission point.

In Eq. (1), L_{AEijk} is the averaged single event level from the aircraft type j on the flight route k at a specific immission point i. The simulation accounts for the real distance between source and receiver, the real flight paths, the geometric and the atmospheric attenuation, additional attenuation effects near the ground, and the spectral rotationally symmetric directivity of the source. The source models for 60 different aircraft types were derived from extensive acoustic measurements of aircraft movements on site (Krebs *et al.*, 2004). N_{Hjk} in Eq. (1) is the yearly average number of movements of the aircraft type j on the flight route k during the hour H. The 1 h-L_{Aeq} of the specific hour is then the energetic sum of all single event levels calculated at the specific immission point i. The hourly average sound pressure levels across 24 h are the basis for all the given noise exposure values used in the present article (see equations below). Error analysis investigations of the FLULA2 model by Thomann (2007) showed that it is possible to reach an extended uncertainty ($k=2$; $p=95\%$) of the calculated 1 h-L_{Aeq} between about 1.0 and 1.5 dB. However, some systematic deviations from the widespread INM model (Olmstead *et al.*, 2001) were found insofar as that the INM yields systematically lower levels in regions with L_{den} below 55 dB. This fact should be accounted for when comparing calculated levels from different models.

Noise exposure of respondents in this article is—for the most part—given in L_{den} values. The L_{den} is defined in terms of the average levels during day, evening, and night and has been proposed as the new uniform noise metric for the European Union (European Union, 2000). For comparison with other studies, the L_{dn} noise metric as well as the $L_{Aeq,24\ h}$ metric will also be used. The metrics are calculated as follows:

$$L_{Aeq,24\ h} = 10\log\left(\frac{1}{24}\cdot\sum_{H=0}^{23} 10^{0.1\cdot L_{Aeq,H}}\right) = 10\log\left(\sum_{H=0}^{23} 10^{0.1\cdot L_{Aeq,H}}\right) - 10\log(24) \quad (2)$$

$$L_{dn} = 10\log\left(\frac{15}{24}\cdot 10^{0.1\cdot(L_{Aeq,d}+K_d)} + \frac{9}{24}\cdot 10^{0.1\cdot(L_{Aeq,n}+K_n)}\right) = 10\log\left(\sum_{H=7}^{21} 10^{0.1\cdot(L_{Aeq,H}+K_d)} + \sum_{H=22}^{6} 10^{0.1\cdot(L_{Aeq,H}+K_n)}\right) - 10\log(24) \quad (3)$$

$$L_{den} = 10\log\left(\frac{12}{24}\cdot 10^{0.1\cdot L_{Aeq,d}} + \frac{4}{24}\cdot 10^{0.1\cdot(L_{Aeq,e}+K_e)} + \frac{8}{24}\times 10^{0.1\cdot(L_{Aeq,n}+K_n)}\right) = 10\log\left(\sum_{H=7}^{18} 10^{0.1\cdot L_{Aeq,H}} + \sum_{H=19}^{22} 10^{0.1\cdot(L_{Aeq,H}+K_e)} + \sum_{H=23}^{6} 10^{0.1\cdot(L_{Aeq,H}+K_n)}\right) - 10\log(24) \quad (4)$$

where $K_d=0$ dB, $K_e=5$ dB, $K_n=10$ dB, and $L_{Aeq,d}$, $L_{Aeq,e}$, $L_{Aeq,n}$ for the L_{den} equation: A-weighted long term L_{Aeq} for the day (0700–1900 h), evening (1900–2300 h), and night (2300–0700 h) period at the immission point; for the L_{dn} equation: A-weighted long term L_{Aeq} for the day (0700–2200 h), and night (2200–0700 h) period at the immission point.

TABLE I. L_{den} exposure categories and number of the respondents for both rounds of the study (the category value denominates the midpoint of the L_{den} category which contains all cases with levels from −2.5 to +2.49 dB around the midpoint)

L_{den} category	Survey 2001 Number of subjects	Percentage	Survey 2003 Number of subjects	Percentage
30	0	0	142	8
35	19	1	138	8
40	314	17	84	5
45	130	7	165	10
50	503	28	511	30
55	516	29	382	22
60	241	13	209	12
65	72	4	87	5
70	21	1	1	0
Total	1816	100	1719	100

Exposure calculations for the survey 2003 were carried out on the basis of the flight operations (movements, fleet mix, and flight paths) during the year 2003. However, the calculation basis for the exposure assessment for the survey 2001 was the flight operations during 2000. The reason for this is that until September 11 in 2001, which was followed by a considerable decrease of air traffic worldwide, there were no relevant differences in exposure per resident between the year 2000 and the year 2001. Because flight routes have not changed either, noise exposure was comparable between the year 2000 and the year 2001 until about end of August. In addition to civil aircraft noise, a part of the respondents from both surveys experienced military jet noise from a military airfield of the Swiss Air Force located about 10 km southeast of the airport. In order to avoid a possible influence of the presence of—or occlusion through—military noise sources on the judgement of civil aircraft noise annoyance, the respondents (288 in the year 2001 and 269 in the year 2003) that experienced average military jet noise in excess of 50 dB ($L_{Aeq,06:00–18:00\ h}$) were excluded in the exposure-effect calculations that are presented in Sec. III B [as regards Fig. 1, and Eqs. (5) and (6)] and Sec. III C [as regards Table III model (C), Eq. (7), and Fig. 4].

C. Sampling procedure

As the (originally) more descriptive objective of the study was to obtain a *representative* picture of the extent of the aircraft noise annoyance problem in a larger population of people around the airport, the sampling strategy did not primarily aim at collecting data for as accurate as possible constructing exposure-effect relationships. Therefore, a quite large proportion of the sample only experienced rather moderate levels of aircraft noise (see Table I). Potential restrictions concerning interpretability of the study results are discussed in Sec. IV A.

1. 2001 survey

The local authorities of all 68 municipalities that were geographically located within a radius of 20 km around the airport were requested to provide census and address information of their inhabitants. From the initial 68 municipalities enquired, a total of 56 agreed to relinquish their address data to us, provided the data were used solely for the intended research purpose and destroyed afterwards. For the 2001 survey, in each municipality, a random sample of 60 inhabitants was drawn which have been living at their current homes for at least one year and who were in the age range from 18 years and up, giving a sample size of 3360. It has to be emphasized that, unlike is the case with address-based sampling techniques, we obtained a random sample of *individuals*, which means the sample is a true random sample of the municipality population and not just a random sample of available addresses. In the year 2001, from the 3360 forwarded questionnaires, 1816 were returned which amounts to a response rate of 54%. The lowest response rate of a municipality was 32%, and the highest was 90%. Aircraft noise L_{den} levels ranged from 35 to 70 dB(A) in the sample. No significant correlation was found between the response rate of a municipality and the average L_{den} of its respondents.

2. 2003 survey

For the 2003 survey, 747 respondents from the 2001 survey were again asked to participate in the 2003 survey because they had agreed that their addresses were kept with us for possible future surveys. This pool of 747 addresses was replenished up to a total of 3360 addresses. As in 2001, nonresponders were reminded after about a month after mailing the questionnaire. 1209 questionnaires were sent back in total (giving a response rate of 36%). The lowest response rate of a municipality was 20% and the highest was 53%. In 2003 we additionally conducted 510 telephone interviews—30 per municipality—in 17 selected municipalities in different areas around the airport, using the same questionnaire. In total, 1719 records were collected for the survey 2003 (510 telephone interviews, 394 reinterviewed, and 815 new respondents). Aircraft noise L_{den} levels ranged from 29 to 68 dB(A) in the sample. No significant correlation was found between the response rate for a municipality and the average L_{den} of its respondents.

III. RESULTS

A. Descriptives

In the year 2001 survey, 49.6% of the respondents were female, and 50.4% were male. The age of the respondents covered a range between 18 and 98, and the average was 47.37 years. The age class distribution was as follows: younger than 20 years: 4%, 20–40: 33%, 40–60: 41%, and older than 60 years: 22%. In the year 2003 survey, 42.2% of the respondents were female and 57.8% were male. The age of the respondents covered a range between 18 and 95, and the average was 50.89 years. The age class distribution was as follows: younger than 20 years: 1%, 20–40: 28%, 40–60: 41%, and older than 60 years: 30%.

Table I displays an overview over the number of respondents in each L_{den} exposure category for two rounds of study.

As can be derived from Table I, a quite large proportion of respondents were affected by only moderate aircraft noise

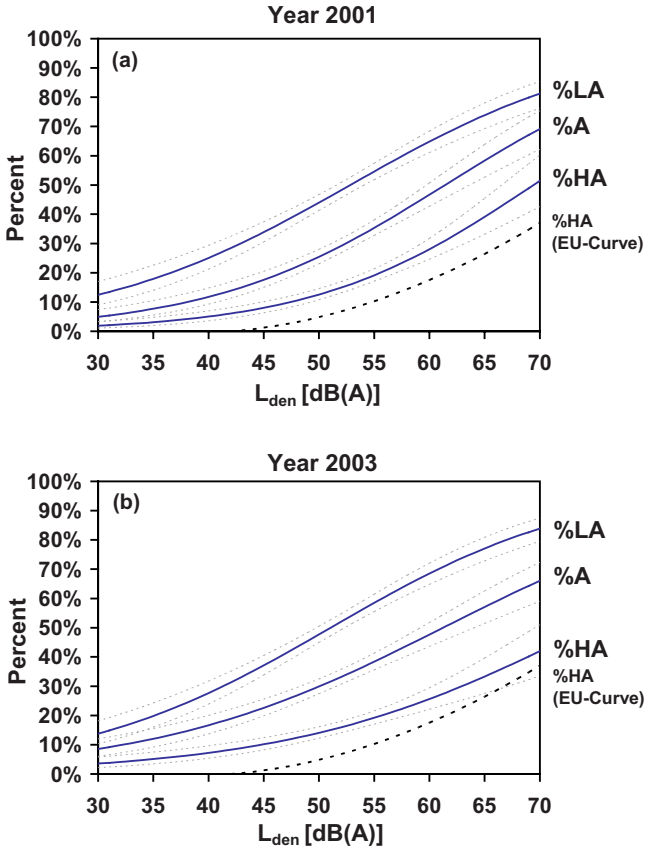


FIG. 1. (Color online) Percent little annoyed (%LA), annoyed (%A), and highly annoyed (%HA) as a function of L_{den}, with 95% confidence limits. (a) Survey 2001, $N=1538$. (b) Survey 2003, $N=1452$. The EU-curve (European Commission, 2002) for %HA is shown for comparison.

exposure as expressed in the L_{den}. This is because the respective sample is a random sample drawn from an area with almost 400 000 inhabitants.

B. Exposure-effect relationships in the years 2001 and 2003

To determine the relationship between noise exposure metrics and the degree of annoyance, we modeled the exposure-effect relationship for the percentage of "little annoyed" (%LA), "annoyed" (%A), and "highly annoyed" (%HA) persons using logistic regression analysis. %HA is the most widely used indicator whereas the %LA and %A indicators better reflect annoyance at rather low exposure levels. The value of the variable HA was defined as 1 if the answer on the annoyance scale (0–10 points) of the questionnaire was greater or equal 8 points. Variable A was coded as 1 when it was greater or equal 6 points, and variable LA when it was greater or equal 4 points. The regression calculations were made with SAS PROC LOGISTIC (SAS version 9, SAS Institute, Cary, NC). Figure 1 shows the curve estimators from the logistic regressions with the L_{den} noise metric, together with their 95% confidence intervals and—for the purpose of comparison—the EU curve (European Commission, 2002).

The percentages of persons expressing different grades of annoyance can be estimated using the logistic functions given in Eqs. (5) (L_{den}) and (6) (L_{dn}).

$$\%\mathrm{HA}_{2001} = 100 \cdot [1 + \exp(-[0.10 \cdot L_{den} - 6.93])^{-1}]$$

$$\%\mathrm{HA}_{2003} = 100 \cdot [1 + \exp(-[0.07 \cdot L_{den} - 5.52])^{-1}],$$

$$\%\mathrm{A}_{2001} = 100 \cdot [1 + \exp(-[0.09 \cdot L_{den} - 5.78])^{-1}]$$

$$\%\mathrm{A}_{2003} = 100 \cdot [1 + \exp(-[0.08 \cdot L_{den} - 4.64])^{-1}],$$

$$\%\mathrm{LA}_{2001} = 100 \cdot [1 + \exp(-[0.08 \cdot L_{den} - 4.50])^{-1}]$$

$$\%\mathrm{LA}_{2003} = 100 \cdot [1 + \exp(-[0.09 \cdot L_{den} - 4.44])^{-1}], \quad (5)$$

$$\%\mathrm{HA}_{2001} = 100 \cdot [1 + \exp(-[0.10 \cdot L_{dn} - 6.93])^{-1}]$$

$$\%\mathrm{HA}_{2003} = 100 \cdot [1 + \exp(-[0.07 \cdot L_{dn} - 5.29])^{-1}],$$

$$\%\mathrm{A}_{2001} = 100 \cdot [1 + \exp(-[0.10 \cdot L_{dn} - 6.03])^{-1}]$$

$$\%\mathrm{A}_{2003} = 100 \cdot [1 + \exp(-[0.07 \cdot L_{dn} - 4.54])^{-1}],$$

$$\%\mathrm{LA}_{2001} = 100 \cdot [1 + \exp(-[0.09 \cdot L_{dn} - 4.80])^{-1}]$$

$$\%\mathrm{LA}_{2003} = 100 \cdot [1 + \exp(-[0.08 \cdot L_{dn} - 4.35])^{-1}]. \quad (6)$$

In general, as it becomes evident from both the curves in Fig. 1 and Eqs. (5) and (6), community annoyance is an increasing function of the L_{den}/L_{dn}. The function curves run about parallel with the generalized curves from Miedema and Oudshoorn (2001), including the EU curve for the L_{den} but are shifted toward the left by about 5–10 dB. This shows that the percentage highly annoyed at a given exposure level in Zurich is actually higher than would be predicted by the Miedema and Oudshoorn curves.

Fidell *et al.* (1985) hypothesized that public awareness of a (upcoming) step change may be enough to evoke higher-than-predicted annoyance. After the necessity of installing new flight routes over in part densely populated areas has been announced by the airport authorities in the year before the first survey 2001 took place, the aircraft noise issue in Zurich received considerable attention by the media. After the new routes were (partly) established in fall 2001, the public interest increased even more and aircraft noise became a top issue on the local political agenda after 2001 (and it is still today). Moderating factors other than mere exposure, such as heightened noise awareness through, e.g., vast media coverage, might evoke a shift of the exposure-effect relationship. Therefore, with a data set where the samples from 2001 and 2003 were merged into one data set, an additional logistic regression analysis was carried out to determine whether the year of the survey has had an impact on annoyance, independent of particular changes in the immission situation. Because in 2003, almost 400 respondents were already represented in the 2001 sample, we treated the year as fixed and the respondent as random effect and used SAS's GLIMMIX procedure to model high annoyance as depending on the year of survey, the survey method (mailed questionnaire versus telephone interview), and the L_{den}. The tests of fixed effects did not provide any evidence for an effect of survey method (questionnaire versus telephone in-

terviews) [$F(1)=0.25$; $p=.62$], nor could we find evidence for a year-of-survey effect [$F(1)=.67$; $p=.41$].

C. Effects of a step change in exposure between 2001 and 2003

Abrupt changes in noise exposure usually produce a change in reaction that is much greater than would be predicted from the change in sound exposure with exposure-effect functions that assume stable conditions (Guski, 2001; Van Kempen and Van Kamp, 2005). In general, when noise exposure increases, the level of annoyance increases to a level that is well beyond than would be expected if the new noise level had existed for a long time. The parallel effect exists for reductions in noise exposure, such that the reaction to the reduced noise is less than would be expected if the new low level had existed for many years. These effects are generally termed "over-reaction effects" to reflect the exaggerated reaction to changes in noise exposure in either direction. Raw and Griffiths (1990) claimed that over-reaction effects can be seen up to nine years after the change.

In the current study we aimed at determining the size range of such an over-reaction effect by taking advantage of the change in the operating plan of the airport that took place between the years the two surveys were carried out. It appears necessary to operationalize change in terms of either an increase or a decrease of number of operations or as a change of average sound exposure which is beyond the amount of normal fluctuations usually observable at airports (Horonjeff and Robert, 1997). The aircraft noise exposure situation around Zurich Airport at the beginning of this decade was characterized by a stepwise adaptation of the operating plan during shoulder hours between October 2001 and October 2003 which, for particular communities, resulted in a substantial change of the shoulder hour exposure. Until October 2001, about 95% of all inbound air traffic to Zurich used German airspace for the final approach on either Runway 14 or 16 (see Fig. 3), the German border being just about 15 km north of the runway ends. As a consequence of newly impeded restrictions pertaining to the usage of south German airspace which were put into force by the German government as of October 19, 2001 (two months after the first survey), the operating plan of the airport was adjusted in basically three subsequent steps that happened during the following one and a half years. By May 2003 almost all inbound flights during 6:00–7:00 h and 21:00–24:00 h on weekdays and between 6:00–9:00 h and 20:00–24:00 h on weekends were relocated from the former northern approach route to landing on Runway 28, now overflying eastern communities not previously affected by aircraft landings. The largest step change in the number of movements occurred after April 2003 when a new instrument landing system (ILS) for Runway 28 begins its operation. The number of shoulder hour landing approaches (06–09 h and 21–23 h) onto that runway during the traffic intense months of June, July, and August raised from 377 in 2001, to 919 in 2002, and to 4748 in 2003. Figure 2 shows the development of aircraft movements on the approach path onto Runway 28 (overflying eastern communities of Kloten, Bassersdorf, Nürensdorf, Lindau, Illnau-Effretikon, Weisslingen, and Zell) during shoulder hours in the years 2001, 2002, and 2003.

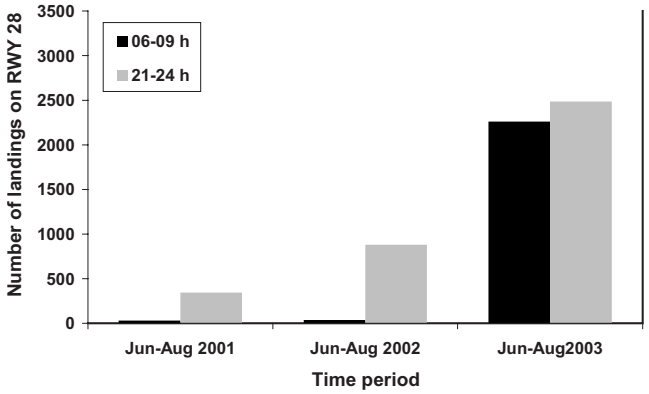


FIG. 2. Number of aircraft landings onto Runway 28 during the months of June, July, and August, in the years 2001, 2002, and 2003.

Considering the fact that the total number of operations at Zurich Airport actually *decreased* after September 2001, the more than tenfold local *increase* in the numbers of landing flights in the morning affecting the east can, for this particular region, be considered a substantial step change. Of course, the changes introduced between 2001 and 2003 as well as the general decline of air traffic in the same time period resulted in a decrease of exposure in many other communities, particularly north of the airport. Figure 3 displays the difference of the daily average $L_{Aeq,06–22\ h}$ between the year 2000 and the year 2003 as a grayscale gradient drawn on a map of the area around the airport.

The magnitude of a (step) change can be conceptualized as exposure difference (in the sense of "current exposure" minus "previous exposure" or "new exposure" minus "old exposure") and be used as a covariate in a exposure-effect relationship model. As concerns the measure to express this change, differences of *average levels* best seem to serve the purpose of forecasting annoyance after changes at airports because in most instances (hourly) average levels are readily available, at least more so than, e.g., a number of aircraft movements above a certain threshold or other means of expressing exposure. As mentioned, major changes in exposure to aircraft noise between 2001 and 2003 occurred predominantly during shoulder hours; therefore, to create a variable of change for each respondent in the subsample between 2001 and 2003, we calculated the difference of the L_{Aeq} for the time period from 6:00 to 9:00 h and the time period from 21:00 to 24:00 h.

In the current study, related records for both surveys, which allow to derive the change effect, were available for the group of respondents that were interviewed twice ($N=394$). The largest differences in exposure between the years 2001 and 2003 in this subsample appeared in the early morning between 6:00 and 7:00 h with the highest 1 h-L_{Aeq} level difference of a respondent at +20.3 dB. Table II shows the frequency distribution of the amounts of change in the subsample between 2001 and 2003. As is evident from Table II, the exposure in 2003 decreased for a majority of respondents, mostly because of the demise of the Swissair company

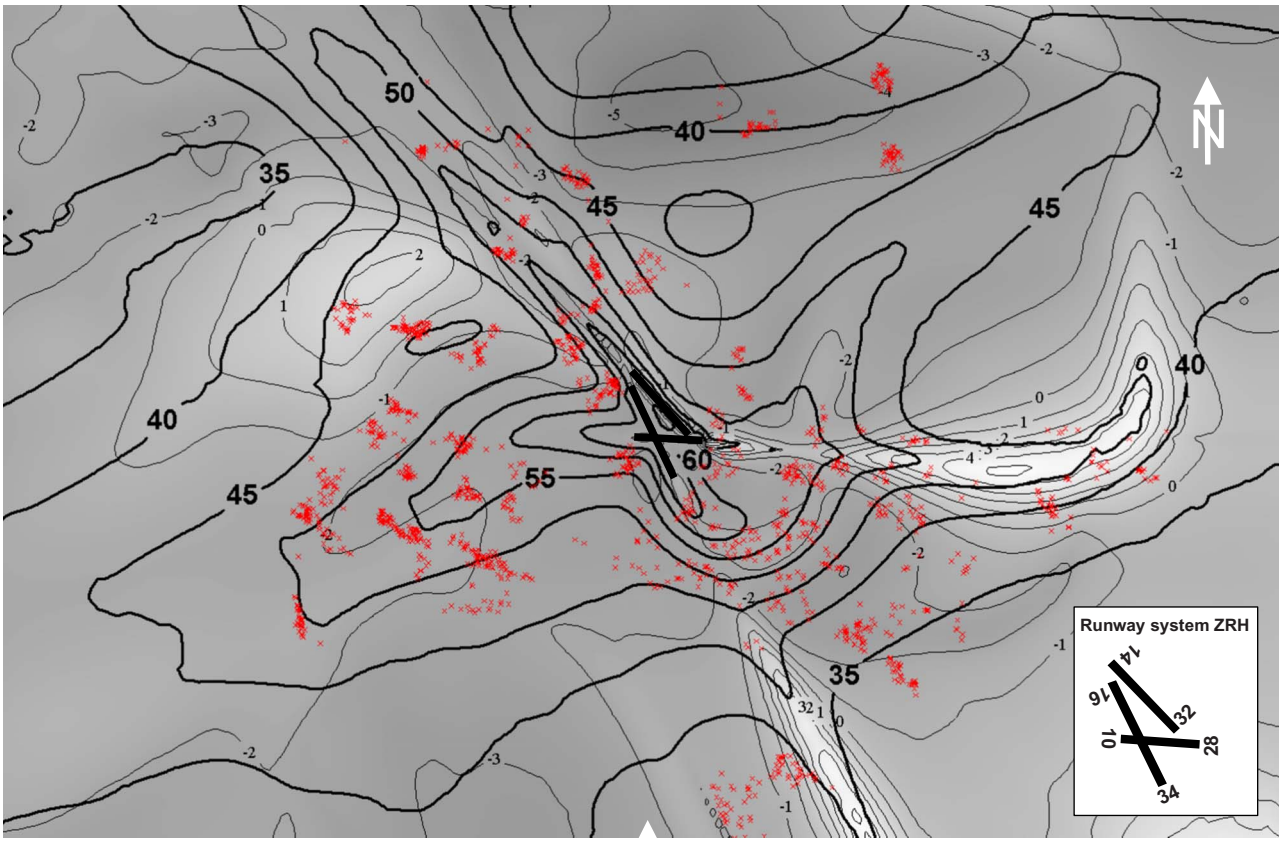


FIG. 3. (Color online) Aircraft noise exposure differences around Zurich Airport in 1 dB steps between the average daily exposure in the year 2003 and the average daily exposure in the year 2000. Negative differences stand for a higher level in 2000, and positive differences for a higher level in 2003. The grayscale gradient reaches from the lowest (dark gray=exposure decrease) to the highest differences (white=exposure increase). The bold curves show the exposure contours of the $L_{Aeq,06-22\ h}$ for the year 2003. The crosses correspond to the places of living of the respondents from the 2001 survey.

in fall 2001 and probably also due to the general reduction of air traffic in the aftermath of the terrorist attacks of September 11.

Two logistic regression models were specified in order to test the influence of the noise exposure level in 2003, as defined by the L_{den}, and the differences of the noise exposure level between 2001 and 2003 during morning and late evening, as well as the interaction between the two variables, on the presence of high annoyance in 2003. The first model (a) contained the change of the L_{Aeq} in the morning (6:00–9:00 h), and the second model (b) contained the change during late evening (21:00–24:00 h). As the interactions between the L_{den} and the differences were not significant in both models, the interaction term was eliminated from both models.

In addition to these two models, we calculated an "average" model (c) for the probability of high annoyance using all responses from both years from people whose place of residence did not exceed 50 dB(A) $L_{Aeq,6-18\ h}$ of military jet noise exposure. For the subset of people meeting this criterion and who were interviewed twice (N=331), the exposure

TABLE II. Frequency distribution of the exposure difference during shoulder hours in the subsample of twice-interviewed respondents.

	Morning (6:00–9:00 h)			Evening (21:–24:00 h)		
Exposure difference (2003–2001)	Count	Percent	Cumulative percent	Count	Percent	Cumulative percent
≥12 dB	0	0	0	0	0	0
≥9 dB	9	2	2	4	1	1
≥6 dB	15	4	6	31	8	9
≥3 dB	18	13	24	52	13	24
≥0 dB	51	13	24	52	13	24
≥−3 dB	161	41	64	178	45	70
≥−6 dB	83	21	86	77	20	89
≥−9 dB	34	9	94	39	10	99
≥−12 dB	23	6	100	4	1	100

TABLE III. Parameter estimates for three logistic regression models modelling the probability of high annoyance

Model		Parameter	Estimate	SE	p	95% Lower CI	95% Upper CI
(a) Morning (N=394)		Intercept	−4.53	1.31	0.00055	−7.10	−1.96
		L_{den}	0.05	0.03	0.00003	0.01	0.1
		Change of $L_{Aeq,06:00–09:00\ h}$	0.16	0.04	0.02830	0.08	0.23
(b) Evening (N=394)		Intercept	−5.32	1.31	0.00005	−7.9	−2.75
		L_{den}	0.07	0.03	0.00554	0.02	0.12
		Change of $L_{Aeq,21:00–24:00\ h}$	0.16	0.04	0.00002	0.09	0.24
(c) Average (N=2685)	(L_{den})	Intercept	−6.10	0.44	<0.00001	−6.96	−5.25
		L_{den}	0.09	0.008	<0.00001	0.07	0.1
	(L_{dn})	Intercept	−6.12	0.45	<0.00001	−7	−5.24
		L_{dn}	0.08	0.008	<0.00001	0.06	0.1
	($L_{Aeq,24\ h}$)	Intercept	−5.33	0.39	<0.00001	−6.09	−4.57
		$L_{Aeq,24\ h}$	0.08	0.008	<0.00001	0.06	0.09

was calculated as the energetic, and the annoyance (on the 11-point scale) as the arithmetic mean of both years. The response code 1 (indicating high annoyance) was assigned to scale values greater or equal 8. The data set for the average model (c) consisted of 2685 individual cases (1230 from the year 2001, 1124 from the year 2003, and 331 from both years). Three logistic regression models were calculated with L_{den}, L_{dn}, and $L_{Aeq,24\ h}$ as predictors. Table III displays the parameter estimates for these models.

Presently, the three average models in Table III (c) covering the L_{den}, L_{dn}, and $L_{Aeq,24\ h}$ noise metric represent the best available estimation of the exposure-effect relationship between aircraft noise and high annoyance for the population around Zurich Airport. The following polynomial approximations of the three models are easier to use and sufficiently accurate in the range between 30 and 70 dB $L_{den}/L_{dn}/L_{Aeq,24\ h}$:

$$\%\,\mathrm{HA}(L_{den}) = (0.0758 - 0.0027 \cdot L_{den} - 2.196 \times 10^{-5} \cdot L_{den}^2 + 1.9982 \times 10^{-6} \cdot L_{den}^3) \cdot 100,$$

$$\%\,\mathrm{HA}(L_{dn}) = (0.0498 - 0.0009 \cdot L_{dn} - 5.9509 \times 10^{-5} \cdot L_{dn}^2 + 2.1983 \times 10^{-6} \cdot L_{dn}^3) \cdot 100,$$

$$\%\,\mathrm{HA}(L_{Aeq,24\ h}) = (0.3011 - 0.0193 \cdot L_{Aeq,24\ h} + 0.0004 \cdot L_{Aeq,24\ h}^2 - 9.5435 \times 10^{-7} \cdot L_{Aeq,24\ h}^3) \cdot 100. \quad (7)$$

Figure 4 shows the exposure-effect curves as predicted with the models (a) and (b) from Table III. The curves for a 5 dB change are drawn as an arbitrary example. They can be regarded as an estimate of the "current" (2003) exposure-effect relationship for the part of the population which experienced a 5 dB change during shoulder hours since 2001. The "no change" curve reflects the estimation of the exposure-effect relationship for the steady state condition in the sample. For an L_{den} exposure in the 50–70 dB range, a recent 5 dB step increase during shoulder hours results in about 15%–20% more highly annoyed persons than would be predicted for the same L_{den} but under steady state conditions.

The curves in both graphs in Fig. 4 show the shift of the exposure-effect-curve when a recent change of exposure in either the morning or evening hours comes into play. The shift of the curve on the X-axis per decibel of change is

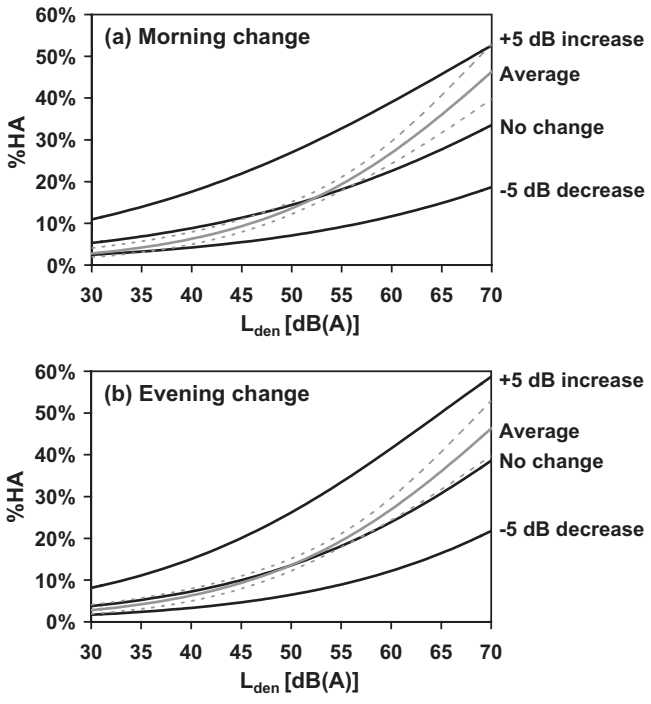


FIG. 4. Estimated curves of the percentage of highly annoyed respondents (%HA) in the year 2003 which were also interviewed in 2001 as a function of L_{den} in 2003. Curves are shown for unchanged exposure (0 dB), a 5 dB step increase, and a 5 dB step decrease of exposure according to the parameters from Table III. In addition, the curve for the average model including 95% confidence boundaries is shown. (a) Morning change (change of $L_{Aeq,06–09\ h}$). (b) Evening change (change of $L_{Aeq,06–09\ h}$)

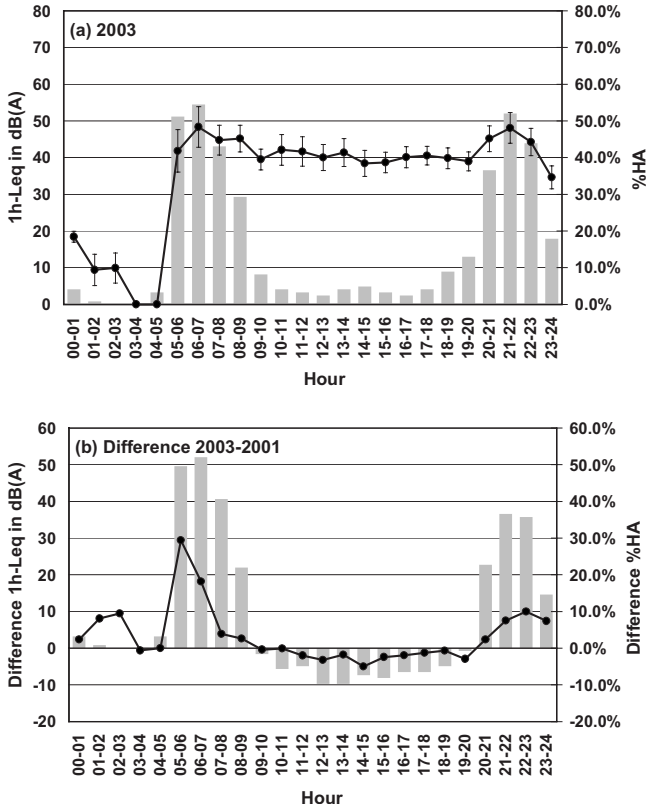


FIG. 5. (a) Percentage of highly annoyed respondents per hour of day (right axis; columns) in four municipalities east of the airport along with the arithmetic average 1 h-L_{Aeq} value (left axis; line with round markers including standard deviation error bars) in the year 2003. (b) Difference of the percentage of highly annoyed respondents in the same group of municipalities, per hour of day, and corresponding difference of the 1 h-L_{Aeq} between the years 2003 and 2001.

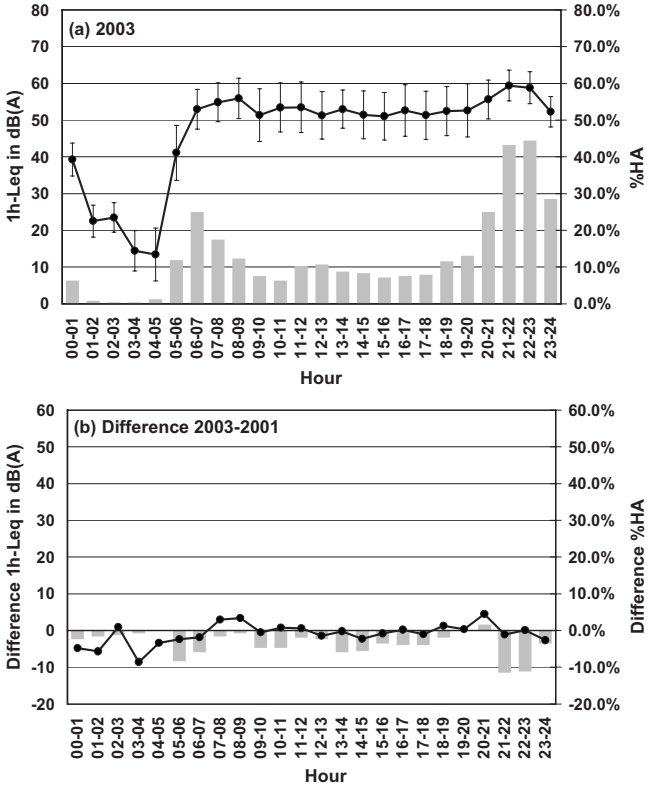


FIG. 6. (a) Percentage of highly annoyed respondents per hour of day (right axis; columns) in seven municipalities north of the airport along with the arithmetic average 1 h-L_{Aeq} value (left axis; line with round markers including standard deviation error bars) in the year 2003. (b) Difference of the percentage of highly annoyed respondents in the same group of municipalities, per hour of day, and corresponding difference of the 1 h-L_{Aeq} between the years 2003 and 2001.

equivalent to the ratio of the estimates for the change effect and the L_{den} effect as given in Table III, i.e., for the morning model, 2.87 dB, and for the evening model, 2.35 dB. Depending on the time of day for which one wants to model the impact of change, one might use either model (a) (if the change pertains to morning hours) or (b) (if the change pertains to evening hours) to predict annoyance after a change in exposure.

D. Changing diurnal variation of annoyance as result of a step change in exposure during shoulder hours

Due to the fact that activities vary across the day, it is plausible to assume that people's responses to noise immissions differ depending on the time/hour of day a particular amount of exposure affects them. Retrospective data on particularly annoying times of day can be collected by asking respondents to mark particular hours of high annoyance in an array of hours of day presented in a questionnaire or in similar form. We investigated diurnal annoyance by asking the respondents for particular times during the day, when they felt highly annoyed in an open question. Up to three time periods could be indicated. The English translation of that particular item in the questionnaire read as follows:

"Are you being particularly highly annoyed (whatever day in the week) by aircraft noise at certain times?" If yes, please indicate these time periods in a 00–24 h format. ( ) no ( ) yes, from:…..to:…..h
from:…..to:…..h
from:…..to:…..h

The data from that item were used to determine from each respondent, for each hour, the occurrence of "high annoyance" which was coded as either "0" (no particular annoyance) or "1" (highly annoyed at that particular hour). We chose this form of questioning mainly for reasons of space (on the questionnaire). It does not adhere to the ICBEN recommendations of gathering annoyance data for deriving exposure-effect relationships, so any binary high annoyance statement of the respondents each hour may not reflect the same degree of annoyance, which would be found with the recommended wording and scales from Fields *et al.*, (2001). However, it is reasonable to assume that both forms of questioning correlate. Respondents not indicating any particular time of high annoyance, which means answering the above question with "no," were still considered valid cases that were not highly annoyed at any time of day.

In Figs. 5(a) and 6(a) the diurnal annoyance profiles, expressed as percentage of respondents claiming to be highly annoyed as determined with the above question, along with the corresponding 1 h-L_{Aeq} values of the decisively new af-

fected communities with increasing exposure during late evening and early morning (municipalities of Illnau-Effretikon, Lindau, Weisslingen, and Zell; $N=154$) are compared with seven communities to the north (municipalities of Bülach, Glattfelden, Hochfelden, Höri, Niederglatt, Oberglatt, and Stadel; $N=238$) which benefited from the new operating plan in terms of slightly decreasing exposure.

Respondents from the communities to the east (Fig. 5) which received "new" noise from landing aircraft during shoulder hours express a rather clear over-reaction in the year 2003. Their degree of annoyance during shoulder hours, as compared to the communities to the north (Fig. 6), which are affected by even higher exposure values (mostly from takeoffs), is higher. During early morning between 5:00 and 7:00 h, almost all annoyance reported in this group is new because the percentage difference as compared to 2001 [Fig. 5(b)] is about the same as the actual annoyance in 2003. The relocation of incoming flights to the eastern region is reflected in decreasing annoyance in the northern communities [Fig. 6(b)] in the hours of 5:00–7:00 and 21:00–23:00.

The data demonstrate in a descriptive fashion that a quite strong over-reaction effect exists in the eastern communities which seems to be specific to particular times during the day.

IV. CONCLUDING REMARKS

A. Study design and uncertainty

The sampling strategy of the current study did not primarily aim at collecting data for as accurate as possible constructing exposure-effect relationships but rather to obtain a representative picture of aircraft noise annoyance incidences in a larger population around the airport. Therefore, no systematic exposure level stratification was aimed for and consequentially, a quite large proportion of the sample only experienced rather moderate levels of aircraft noise (see Table I). This could be considered a weakness of the design. Since low levels are generally more afflicted with uncertainty of the acoustic calculations, impact assessment at lower acoustic levels is more uncertain too. However, we have evidence that also low levels of noise can still elicit annoyance reactions and, given the relatively large numbers of people experiencing these kinds of low and moderate levels around many airports, surely deserve study. The combined sample size of 3535 respondents over the two survey years was comparatively large.

The confidence intervals of the exposure-effect functions in Figs. 1 and 4 and the standard errors reported in Table III pertain to the uncertainty of the annoyance measurement in the sample (the Y-axis) not the uncertainty of exposure calculations (the X-axis). Therefore, the "physically true" confidence boundaries are most probably wider. We did not calculate the extended uncertainty of the exposure-effect relationship since the nonconsideration of acoustic uncertainty is a shortcoming of almost all annoyance studies of this kind. Accounting for acoustic uncertainty would possibly compromise the comparability of the confidence intervals in this study with the ones from other studies.

B. Exposure-effect relationships

The two surveys carried out around Zurich Airport in 2001 and 2003 provide the most up to date exposure-effect functions for aircraft noise in Switzerland. The L_{den} function curve for the average model [Eq. (7)] runs about parallel with the generalized EU curve published in the EU position paper on noise annoyance (European Commission, 2002) but is shifted toward the left by about 5 to 10 dB, indicating that the percentage of highly annoyed persons is actually higher than would be predicted by the EU curve. The current data provide additional evidence that annoyance has increased in the past decades and that aircraft noise annoyance of residents in Europe (and probably elsewhere too) nowadays no longer seems to be well reflected in the EU curve, confirming other recent findings from the UK, Germany, The Netherlands, Greece, Spain, and Italy (Babisch *et al.*, 2007; Breugelmans *et al.*, 2004; Department for Transport, 2007; Schreckenberg and Meis, 2006). Furthermore, we could not determine any threshold exposure value below which %HA would be zero. That means that the feeling of being highly annoyed by aircraft noise in our sample was not necessarily bound to acoustic exposure at all.

Respondents in our study are slightly less annoyed—especially at exposure values above about 50 dB(A) L_{dn}—than in comparable recent European studies such as from Amsterdam 2002 (Breugelmans *et al.*, 2005) or Frankfurt 2005 (Schreckenberg and Meis, 2006). A possible explanation for this difference could be the smaller number of aircraft operations in Zurich, as compared to large European hub airports that are particularly often in the focus of annoyance studies (London Heathrow, Amsterdam, and Frankfurt). An alternative explanation would also be that the observed differences in annoyance are an effect of exposure calculation differences of the different aircraft noise models used in the respective studies. As it has been mentioned, the Swiss calculation model FLULA2 is suspected to yield somewhat higher exposure estimates at calculated levels below 55 dB(A) as, e.g., the INM. In this light, the comparison between the Zurich data and the EU curve in Fig. 1 might be interpreted with caution also. However, only a systematical error analysis that accounts for uncertainty of the calculated levels and uncertainty of the parameters of the exposure-effect model could clarify this issue.

C. Influence of an exposure change on annoyance

In the past, several authors emphasized the importance of evaluating noise effects due to changes in exposure (Fidell *et al.*, 2002; Fields *et al.*, 2000; Guski, 2003; Guski, 2004; Horonjeff and Robert, 1997; Job, 1988a; Krog and Engdahl, 2004). The study we reported about was to a certain degree comparable to the step change investigation carried out at Vancouver International Airport (Fidell *et al.*, 2002) but is characterized by a different approach to data analysis. For the forecasting of noise annoyance after operational changes, the ability to predict the annoyance shift as a function of the amount of change is very useful. Mainly for reasons of usability, we describe a model that operationalizes the change as a decibel difference value on a continuous scale reaching

from a negative difference (decrease of exposure) to a positive one (increase of exposure). Although tiny increases or decreases of L_{Aeq} values might not be perceivable on an individual level, we could supply evidence that previously experienced exposure change significantly shifts the exposure-annoyance relationship of residents—at least for several months. Therefore, it seems justified to promote a model that incorporates that change and expresses it in a manner that airport authorities or communal agencies can handle, namely, decibel values and decibel differences. In contrast to generalized exposure-effect curves, which do not account for change effects, the models (a) and (b) in Table III use both the *current exposure* and the *change of exposure* as predictors of high annoyance. Therefore, they provide a useful addition to other exposure-effect models in the literature (e.g., the ones from Miedema and Vos, 1998). Because in the current case, the operational change at Zurich Airport was limited to the early morning and the late evening hours, we calculated two separate models, one for the impact of the change in the evening, and one for the impact of the change in the early morning. From the data that were available to us, we could not draw any conclusions pertaining to the effects of an overall general step change that affects the whole day or night period (e.g., opening of an entirely new airport).

The aircraft noise situation that prevailed during the study period can be described as a series of distinctive changes within a period of roughly one and a half years, whereas the last step that took place after April 17, 2003 (Installation of ILS) was particularly large. The following four to five months before the second survey obviously were not long enough a period for the over-reaction effect to diminish, supporting the assumption that the duration of over-reaction effects is rather a question of months or years than days or weeks.

The average exposure in the year 2003, which was the basis for the exposure difference calculations, was affected by two major operational changes: After April 17, 2003, a large part of inbound shoulder hour flights previously using another runway were relocated to use Runway 28 for landing. The landing regime was changed again after October 30, 2003, after the second survey: Inbound flights in the early morning now used Runway 34 and approached the airport from the south (see grayscale gradient in Fig. 3), releasing the east from the early morning noise burden. Therefore, the reported morning exposure as well as its changes are somewhat smaller than the average exposure that prevailed during the five months before the second survey (August–September 2003) in the eastern communities. Thus the estimation of the change effect as pertaining the morning shoulder hours can be termed rather conservative.

It has been hypothesized (Fidell *et al.*, 1985; Van Kempen and Van Kamp, 2005) that public awareness or the expectancy of an upcoming change alone may be enough to evoke an advanced over-reaction so that higher-than-predicted annoyance would be reported by respondents even before an exposure change actually takes place. Although the upcoming changes in the flight regime were announced in the media for more than a year in advance, we found no evidence that the eastern residents prematurely "reacted" to the future noise situation. If that would have been the case, a probably much smaller difference of high annoyance or no difference at all during shoulder hours would have been found between 2001 and 2003 in this population group. Therefore, we conclude that residents actually rate their annoyance based on real experienced exposure and not on any imaginary future noise scenario. This conclusion is further supported by the fact that we could not find any statistical evidence that the year of survey had an independent effect on the probability of high annoyance in the whole sample despite the fact that the aircraft noise issue in 2003 was much more intense a matter of public debate than in 2001.

ACKNOWLEDGMENTS

This study was funded by the Swiss Federal Office for the Environment, the Swiss Federal Office of Public Health and Unique Flughafen Zürich AG. We are grateful to these institutions for their financial support. We would further like to thank Joachim Kastka, Tommaso Meloni, Walter Krebs, Philipp Huber, Helmut Krueger, Tommy Läubli, Hans-Rudolf Roth, and Werner Stahel who contributed to improving the study design, to data collection and processing, and data analysis. All authors declare no conflicts of interest regarding the work presented in this article.

Arbeitsgemeinschaft für sozio-psychologische Fluglärmuntersuchungen. (**1974**). "Soziopsychologische Fluglärmuntersuchung im Gebiet der drei Schweizer Flughäfen Zürich, Genf, Basel (Socio-psychological investigation of aircraft noise around the tree Swiss airports Zurich, Geneva, Basel)," Eidgenössisches Luftamt, Bern.

Babisch, W., Houthuijs, D., Pershagen, G., Cadum, E., Velonakis, M., Katsouyanni, K., and Jarup, L. (**2007**). "Associations between road traffic noise, aircraft noise and noise annoyance. Preliminary results of the HYENA study," Paper presented at the 19th International Congress on Acoustics ICA, Madrid, Spain.

Beck, U. (**1992**). *Risk society: Towards a New Modernity* (Sage, London).

Berglund, B., and Lindvall, T. (**1995**). "Community noise," *Archives of the Centre for Sensory Research* (Stockholm University and Karolinska Institute, Stockholm), Vol. **1**, No. 1.

Breugelmans, O. R. P., van Wiechen, C. M. A. G., van Kamp, I., Heisterkamp, S. H., and Houthuijs, D. J. M. (**2004**). "Gezondheid en beleving van de omgevingskwaliteit in de regio Schiphol: 2002 [RIVM rapport 630100001/2004] (Health and quality of life near Amsterdam Schiphol Airport: 2002 Interim report [RIVM report 630100001/2004])," avaliable at: http://www.rivm.nl/bibliotheek/rapporten/630100001.pdf (last viewed on 01/16/2008).

Brink, M., Thomann, G., Huber, P., and Schierz, C. (**2007**). "A new noise impact assessment method for noise policy," Paper presented at the Internoise 2007, Istanbul, Turkey.

Bröer, C. (**2007**). "Aircraft noise and risk politics," Health Risk Soc. **9**, 37–52.

Bröer, C., and Wirth, K. (**2004**). "Mehr Belästigung bei gleichem Pegel. Wieso Flugzeuggeräusche heute möglicherweise lästiger sind als vor 40 Jahren (More annoyance at identical levels. Why aircraft sounds may be more annoying today than 40 years ago)," Zeitschrift für Lärmbekämpfung **51**, 118–121.

Department for Transport. (**2007**). "Attitudes to Noise from Aviation Sources in England (ANASE) Final Report," avaliable at http://www.dft.gov.uk/pgr/aviation/environmentalissues/Anase/ (last viewed on 06/12/2008).

European Commission. (**2002**). "Position paper on dose response relationships between transportation noise and annoyance," avaliable at http://ec.europa.eu/environment/noise/pdf/noise_expert_network.pdf (last viewed on 01/16/2008).

European Union. (**2000**). "Proposal for a European Parliament and Council Directive on the Approximation of the Laws of the Member States Relat-

ing to the Assessment and Reduction of Environmental Noise," EU, Brussels.
Fidell, S., Barber, D. S., and Schultz, T. J. (**1991**). "Updating a Dosage Effect Relationship for the Prevalence of Annoyance Due to General Transportation Noise," J. Acoust. Soc. Am. **89**, 221–233.
Fidell, S., Horonjeff, R., Mills, J., Baldwin, E., Teffeteller, S., and Pearsons, K. (**1985**). "Aircraft noise annoyance at three joint air carrier and general aviation airports," J. Acoust. Soc. Am. **77**, 1054–1068.
Fidell, S., Silvati, L., and Haboly, E. (**2002**). "Social survey of community response to a step change in aircraft noise exposure," J. Acoust. Soc. Am. **111**, 200–209.
Fields, J. M., De Jong, R. G., Gjestland, T., Flindell, I. H., Job, R. F. S., Kurra, S., Lercher, P., Vallet, M., Yano, T., Guski, R., Felscher-Suhr, U., and Schumer, R. (**2001**). "Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation," J. Sound Vib. **242**, 641–679.
Fields, J. M., Ehrlich, G. E., and Zador, P. (**2000**). "Theory and design tools for studies of reactions to abrupt changes in noise exposure [NASA Contractor Report No. CR-2000-210280]," Hampton, VA: NASA Langley Research Center.
Griffiths, I. D., and Raw, G. J. (**1989**). "Adaptation to changes in traffic noise exposure," J. Sound Vib. **132**, 331–336.
Guski, R. (**2001**). "Der Referenten-Entwurf zum Fluglärmgesetz aus der Sicht eines Wirkungsforschers ("The draft bill of aircraft noise regulation as seen by a noise effects researcher")," Zeitschrift für Lärmbekämpfung **48**, 130–131.
Guski, R. (**2003**). "How to predict future annoyance in planning?" Proceedings of the Eighth International Congress on Biological Effects of Noise ICBEN Rotterdam, The Netherlands, pp. 255–256.
Guski, R. (**2004**). "How to forecast community annoyance in planning noisy facilities," Noise Health **6**, 59–64.
Horonjeff, R. D., and Robert, W. E. (**1997**). "Attitudinal responses to changes in noise exposure in residential communities (Report number: HMMH-293350, NAS 1.26:205813, NASACR-1997-205813)," Hampton, VA: National Aeronautics and Space Administration (NASA) Langley Research Center.
Job, R. F. S. (**1988a**). "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction," J. Acoust. Soc. Am. **83**, 991–1001.
Job, R. F. S. (**1988b**). "Overreaction to Changes in Noise Exposure—the Possible Effect of Attitude," J. Sound Vib. **126**, 550–552.
Job, R. F. S., and Hatfield, J. (**2003**). "A model of responses to changes in noise exposure," Proceedings of the Eighth International Congress on Biological Effects of Noise ICBEN Rotterdam, The Netherlands , pp. 232–234.
Krebs, W., Bütikofer, R., Plüss, S., and Thomann, G. (**2004**). "Sound source data for aircraft noise simulation," Acta. Acust. Acust. **90**, 91–100.
Krog, N. H., and Engdahl, B. (**2004**). "Annoyance with aircraft noise in local recreational areas, contingent on changes in exposure and other context variables," J. Acoust. Soc. Am. **116**, 323–333.
Miedema, H., and Oudshoorn, C. (**2001**). "Annoyance from transportation noise: Relationships with exposure metrics DNL and DENL and their confidence intervals," Environ. Health Perspect. **109**, 409–416.
Miedema, H., and Vos, H. (**1998**). "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445.
Oliva, C. (**1998**). "Belastungen der Bevölkerung durch Flug- und Strassenlärm (People's exposure to aircraft and road traffic noise)," (Duncker&Humblot, Berlin).
Olmstead, J. R., Fleming, J. M., and Gulding, J. M. (**2001**). "Integrated Noise Model (INM), Version 6.0, Technical Manual [Report No FAA-AEE-02-01]," Federal Aviation Administration (FAA), Department of Tranportation, Washington DC.
Pietrzko, S. J., and Hofmann, R. F. (**1988**). "Prediction of A-Weighted Aircraft Noise Based on Measured Directivity Patterns," Appl. Acoust. **23**, 29–44.
Raw, G. J., and Griffiths, I. D. (**1990**). "Subjective response to changes in road traffic noise: A model," J. Sound Vib. **141**, 43–54.
Schreckenberg, D., and Meis, M. (**2006**). "Effects of aircraft noise on noise annoyance and quality of life around Frankfurt Airport. Final abbridged report," avaliable at http://www.verkehrslaermwirkung.de/FRA070222.pdf (last viewed on 03/15/2008).
Schultz, T. J. (**1978**). "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405.
Thomann, G. (**2007**). "Mess- und Berechnungsunsicherheit von Fluglärmbelastungen und ihre Konsequenzen (Uncertainties of measured and calculated aircraft noise and consequences in relation to noise limits) [PhD thesis, ETH Zurich]," avaliable at http://e-collection.ethbib.ethz.ch/show?type=diss&nr=17433 (last viewed on 3/20/2008).
van Kempen, E. E., and van Kamp, I. (**2005**). "Annoyance from air traffic noise. Possible trends in exposure-response relationships [Report No 01/2005 MGO EvK]," Dutch Ministry of Housing and Environmental and Spatial Planning, Bilthoven.

A transient boundary element method model of Schroeder diffuser scattering using well mouth impedance

Jonathan A. Hargreaves[a] and Trevor J. Cox
Acoustics Research Centre, The University of Salford, Manchester M5 4WT, United Kingdom

(Received 20 May 2008; revised 31 July 2008; accepted 19 August 2008)

Room acoustic diffusers can be used to treat critical listening environments to improve sound quality. One popular class is Schroeder diffusers, which comprise wells of varying depth separated by thin fins. This paper concerns a new approach to enable the modeling of these complex surfaces in the time domain. Mostly, diffuser scattering is predicted using steady-state single frequency methods. A popular approach is to use a frequency domain boundary element method (BEM) model of a box containing the diffuser, where the mouth of each well is replaced by a compliant surface with appropriate surface impedance. The best way of representing compliant surfaces in time domain prediction models, such as the transient BEM is, however, currently unresolved. A representation based on surface impedance yields convolution kernels which involve future sound, so is not compatible with the current generation of time-marching transient BEM solvers. Consequently, this paper proposes the use of a surface reflection kernel for modeling well behavior and this is tested in a time domain BEM implementation. The new algorithm is verified on two surfaces including a Schroeder diffuser model and accurate results are obtained. It is hoped that this representation may be extended to arbitrary compliant locally reacting materials.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2982420]

PACS number(s): 43.55.Br, 43.20.Px, 43.20.Fn [SFW] Pages: 2942–2951

I. INTRODUCTION AND OVERVIEW

Room acoustic diffusers can be used to treat critical listening environments to improve speech intelligibility and to make music sound better.[1] Such devices are characterized by the uniformity of their scattering which may be measured under anechoic conditions,[2] a time consuming and expensive process. An alternative is to predict this dispersion using a numerical model, and the boundary element method (BEM) is well suited to this task.[3] The speed and low cost of this approach aid prototyping of new designs and even allow automated optimization of treatments to be performed.[4] In a BEM model only the boundaries between obstacles and air are modeled as it is known how sound travels unobstructed. This produces smaller simpler meshes compared to volumetric methods such as finite element and finite difference time domain (FDTD). It also permits an unbounded volume of air to be modeled, making it ideal for free-field scattering scenarios.

Most BEMs assume harmonic excitation so the unknowns are time invariant and complex. While this frequency domain analyses is a useful tool, the transient behavior witnessed in the real world may only be recovered by solving many frequency domain models and then applying an inverse discrete Fourier transform (DFT). An alternative is to drop the time-invariant assumption and formulate the BEM in the time domain as is presented herein. This approach was first published by Friedman and Shaw in 1962,[5] however, its implementation is problematic and consequently the method is still not in widespread use in acoustics. Some of the key issues are outlined below.

A. Time domain BEM stability

The discretized boundary integral equations (BIEs) forming the time domain BEM are typically solved by marching a solution on in time from known initial conditions, usually silence. However, being iterative this process has the potential for instability, a major impediment to the algorithm's widespread use. Rynne[6] observed that similar instabilities affect all time domain BEM models regardless of the application, implying that this behavior is fundamental to the method rather than the problem considered.

The dominant analysis of this phenomenon is by the singularity expansion method.[7] This expresses the system's response to excitation as a sum of resonant poles, each with its own natural frequency and damping. In discrete time, each pole is a complex scalar describing the magnitude and phase change the corresponding mode undergoes in a time-step; hence a mode with a pole of magnitude greater than unity will grow and cause the solver to diverge. These discrete poles are closely related to the eigenvalues of the state-transition process which may be found numerically.[8–10] Such modes should be prohibited by the initial conditions, but in practice they can be seeded by numerical error in the solver; hence the onset of instability can appear highly random and implementation dependent. In addition, Rynne and Smith[7] suggested that pole locations are perturbed by discretization error, so a pole of the BIE which is just stable may lead to a pole of the discretized system which is unstable.

When the problem of sound scattering from a body is stated as a BIE, the restriction that sound cannot travel through the body is lost and a continuation of the exterior medium, in this case, an air-filled cavity, is effectively created inside the body's bounding surface. At certain frequen-

[a]Electronic mail: j.a.hargreaves@salford.ac.uk

 © 2008 Acoustical Society of America

cies, this cavity may resonate, storing energy so the time-invariant frequency domain BEM has a nonunique solution. In the time domain problem, these resonances correspond to oscillatory poles, borderline stable and likely candidates for corruption into divergence by numerical error. Such poles are not physically relevant so their removal is acceptable and improves solver performance. One method that achieves this in the frequency domain is the Burton and Miller formulation,[11] and has been transferred to the time domain as the combined field integral equation (CFIE).[12]

The BEM is a wave based method and its computational cost increases rapidly with frequency. Acceleration algorithms[13,14] have been published to address this issue but, as these are derived from the time-marching solvers for which instability issues remain, the focus herein remains on modeling smaller problems in a nonaccelerated fashion. In addition, some interesting work has been done on alternative solvers[15–19] that may be less sensitive to divergent poles than the current time-marching generation.

B. Modeling Schroeder diffusers

The class of diffuser considered in this paper is the phase grating diffuser, whose development can be traced back to the pioneering work of Schroeder.[20,21] These comprise a series of wells of differing depths according to a number theoretic sequence, separated by thin fins. Sound waves entering each well emerge following the time taken for them to travel to the bottom of the well, reflect, and travel back to the mouth. These delays are optimally decorrelated so the cumulative scattered sound is widely dispersed. Because the wells store sound energy and then reradiate it, the scattered sound is diffused in both space and time; recently, this transient behavior has begun to attract research interest.[22,23] In this paper, a one-dimensional diffuser based on the quadratic residue sequence will be considered; these are designed to diffuse in one plane only and take the form of an extruded cross section as depicted in Fig. 1(a), where the fins are shown partially transparent.

Modeling the two sides of a thin fin aggravates an issue in the BEM known as thin shape breakdown.[24] This may be circumvented by using an open surface BEM which considers the surface to be comprised of thin rigid plates. This formulation has previously been implemented in the time domain;[25] however, it is unsuitable for modeling the solid part of the diffuser as it supports cavity resonances so is prone to instability.

Another modeling approach used in the frequency domain, which avoids these issues, is to model the mouth of each well as a compliant surface; thus the mesh is simplified and becomes a box enclosing the device [Fig. 1(b)]. A well's behavior is described by the surface impedance of its mouth, a quantity ideal for use with BEM, which may be found by assuming that all propagation in the well is axial with negligible losses, so every point at the well mouth reacts locally. At first, this may seem unrealistic, but it has been numerically shown to produce good results,[26] and Schroeder's phase grating model makes these assumptions too. An equivalent time domain model is sought.

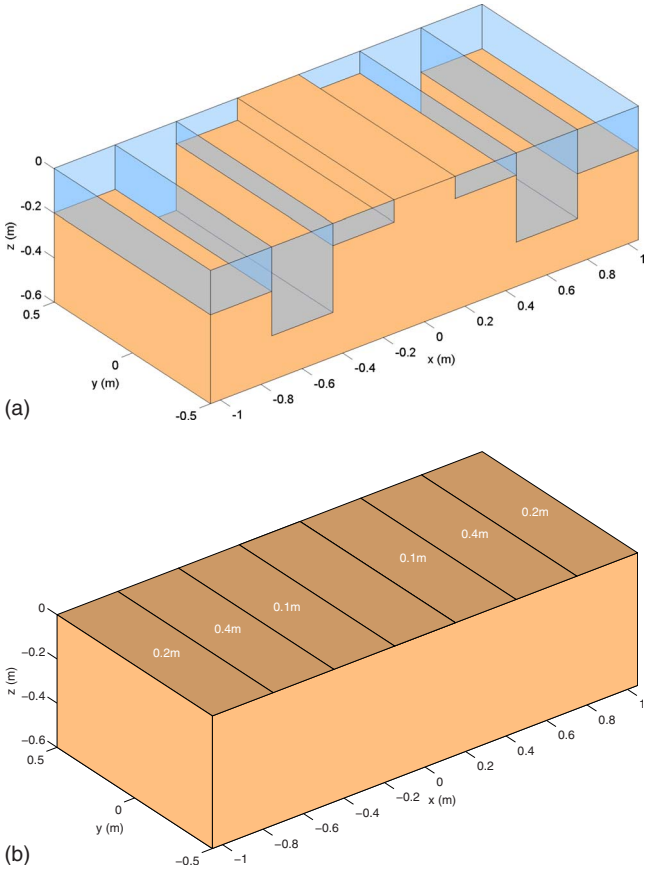


FIG. 1. (Color online) QRD (a) thin plate model, and (b) well mouth impedance model.

Differential boundary conditions may be used to model simple compliant materials such as broadband absorbers[27,28] and limp membranes.[29] However, finding such from arbitrary surface impedance data is more complicated,[30] although Drumm and Lam[31] successfully fitted infinite impulse response filters to surface absorption data. It has been suggested[32] that a convolution between waves traveling perpendicularly into and out of the body may be a more robust approach; this is adopted for the application herein and found to be effective.

This paper is structured as follows: Sec. II introduces the boundary integral formulation of the scattering problem and the CFIE. Section III introduces the new time domain well mouth model and describes its substitution into the BEM. The discretization process and time-marching solver are specified in Sec. IV. Verification results are shown and discussed in Sec. V followed by the conclusions in Sec. VI. Finally, details of the numerical integration procedure are outlined in the Appendix.

II. BOUNDARY INTEGRAL EQUATION FORMULATIONS

Figure 2 depicts a scattering problem, comprising an obstacle submerged in a connected medium Ω_+ with equilibrium density ρ_0 which obeys the linear acoustic wave equation with speed of sound c. S is a surface conformal to the obstacle and sufficiently close that the obstacle's surface properties may be ascribed to it; thus the obstacle resides in the interior domain Ω_-. S_∞ is the extent of the medium. $\mathbf{x}$ and

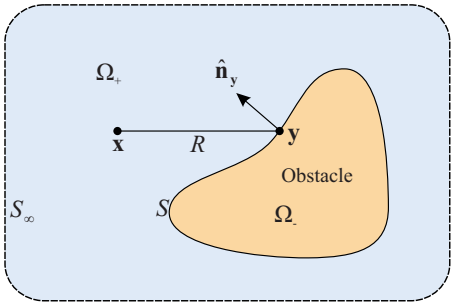


FIG. 2. (Color online) A scattering problem comprising an obstacle submerged in a connected medium. S is a surface conformal to the obstacle; hence the medium is said to be external to S.

$\mathbf{y}$ are the three-dimensional Cartesian vectors defining the observation and radiation points respectively, $R=|\mathbf{x}-\mathbf{y}|$ is the distance between them, and $\hat{\mathbf{n}}_\mathbf{y}$ is the surface normal unit vector at $\mathbf{y}$.

Sound is represented by the velocity potential φ which, while not a physical quantity, has the convenient property that both pressure p and particle velocity $\mathbf{v}$ may be derived from it:

$$p(\mathbf{x},t) = -\rho_0 \dot{\varphi}(\mathbf{x},t), \tag{1}$$

$$\mathbf{v}(\mathbf{x},t) = \nabla\varphi(\mathbf{x},t), \tag{2}$$

where t is the time and a dot above a quantity indicates temporal differentiation. An incident disturbance $\varphi_i(\mathbf{x},t)$ exists in Ω_+ but does not reach the obstacle while $t \leqslant 0$. When $\varphi_i(\mathbf{x},t)$ does reach the obstacle, a wave $\varphi_s(\mathbf{x},t)$ is scattered such that the total disturbance $\varphi_t(\mathbf{x},t)=\varphi_i(\mathbf{x},t)+\varphi_s(\mathbf{x},t)$ matches the surface properties of the obstacle; thus this is an initial-boundary-value problem. Application of Green's theorem[33] allows the propagation of $\varphi_s(\mathbf{x},t)$ in Ω_+ to be stated as the Kirchhoff integral equation (KIE) over its boundary $S \cup S_\infty$. In practice, S_∞ is chosen so distant that its contribution does not arrive within the modeling duration, so the integration domain may be reduced to S. This statement describes scattered velocity potential in a manner equivalent to the classical Huygens principle, being the propagation of the wavefront at each point on the boundary summed together:

$$\varphi_s(\mathbf{x},t) = \iint_S [\varphi_t(\mathbf{y},t)^* \hat{\mathbf{n}}_\mathbf{y} \cdot \nabla_\mathbf{y} g(R,t) - g(R,t)^* \hat{\mathbf{n}}_\mathbf{y} \cdot \nabla_\mathbf{y} \varphi_t(\mathbf{y},t) d\mathbf{y}]. \tag{3}$$

The term $\hat{\mathbf{n}}_\mathbf{y} \cdot \nabla_\mathbf{y} \varphi_t(\mathbf{y},t)$ is the surface normal component of the particle velocity and is termed "normal velocity" for brevity. Despite the relation given in Eq. (2), normal velocity and velocity potential are independent fields on S since surface normal derivatives cannot be found from quantities only known on a surface. * denotes temporal convolution and $g(R,t)$ is the time domain Green's function, which describes how sound travels from an instantaneous point source at $t=0$ to a point observer at distance R, given by

$$g(R,t) = \frac{\delta(t-R/c)}{4\pi R}, \tag{4}$$

where $\delta(\cdots)$ is a Dirac delta function. This delay term in the numerator encapsulates the finite propagation speed of sound, thereby dictating the domain of dependence of the scattered wave in Eq. (3) and ensuring causality in the resulting algorithm.

Consideration of the obstacle's boundary condition at $\mathbf{x}$ allows solution for the total surface sound, from which scattered sound at any desired off-surface point may be evaluated.

Specifically the CFIE is equivalent to the boundary condition $(1-\alpha)p_t(\mathbf{x},t)=\alpha\rho_0 c\hat{\mathbf{n}}_\mathbf{x}\cdot\mathbf{v}_t(\mathbf{x},t)$ when the limit is taken as $\mathbf{x}$ approaches S from the inside. The corresponding integral operator may be expressed as

$$\begin{aligned} L_c\{\varphi_t(\mathbf{x},t)\} &= (1-\alpha)\dot{\varphi}_i(\mathbf{x},t) + \alpha c\hat{\mathbf{n}}_\mathbf{x}\cdot\nabla\varphi_i(\mathbf{x},t) \\ &= -\left[(1-\alpha)\frac{\partial}{\partial t} + \alpha c\hat{\mathbf{n}}_\mathbf{x}\cdot\nabla_\mathbf{x}\right] \\ &\quad \times \iint_S [\varphi_t(\mathbf{y},t)^*\hat{\mathbf{n}}_\mathbf{y}\cdot\nabla_\mathbf{y} g(R,t) \\ &\quad - g(R,t)^*\hat{\mathbf{n}}_\mathbf{y}\cdot\nabla_\mathbf{y}\varphi_t(\mathbf{y},t)]d\mathbf{y} \quad \text{if } \mathbf{x}\in S_-. \end{aligned} \tag{5}$$

This formulation is stated[12] to be the time domain equivalent to the Burton and Miller formulation[11] commonly used in frequency domain acoustic BEMs. It differs slightly from the latter, in particular, with regard to the range of values taken by the real scalar blend parameter α, and matches more closely its namesake in the electromagnetic BEM formulation.[34]

When $\alpha=\frac{1}{2}$ the boundary condition which founds the CFIE simplifies to $p_t(\mathbf{x},t)/\hat{\mathbf{n}}_\mathbf{x}\cdot\mathbf{v}_t(\mathbf{x},t)=\rho_0 c$, a condition satisfied by any plane wave propagating in the direction of $\hat{\mathbf{n}}_\mathbf{x}$, that is out of the cavity. More generally, it has been shown that when $0<\alpha<1$ energy flows out of the cavity and it cannot support resonant modes.[35] Consequentially, L_c has been shown to grant stability superior to alternate operators for a variety of test geometries. Therefore, it is desirable to derive the BEM algorithm from the CFIE for all compatible scattering obstacles.

III. WELL MOUTH SURFACE MODEL

Consider a well of constant cross section and depth d, as depicted in Fig. 3. It is assumed that the width of the well is small with respect to wavelength so that only axial plane wave propagation is supported. Accordingly, the sound in the well may be expressed as the superposition of two plane waves traveling into and out of the well, designated φ_{in} and φ_{out}. The rigid boundary condition at the base of the well dictates that the incoming wave is reflected, such that the outgoing wave is the incoming wave with a fixed time delay.

Under the assumption of time-harmonic oscillation $\varphi(z,t)=\text{Re}\{e^{-i\omega t}\Phi(z,\omega)\}$, the mouth of the well may be characterized by this well known expression for surface impedance at the well mouth $(z=0)$:

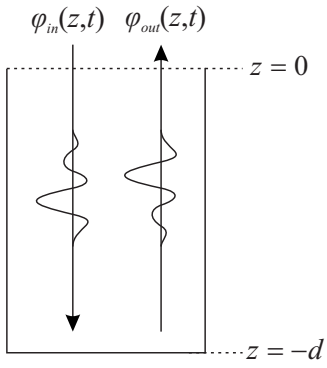


FIG. 3. A simple well with constant cross section, depth d, and mouth at $z=0$.

$$Z(\omega)=\frac{P_t(0,\omega)}{V_{t,\mathrm{in}}(0,\omega)}=i\rho_0 c\cot(kd), \tag{6}$$

where $k=\omega c^{-1}$ is the wavenumber, $P_t(z,\omega)$ is the total pressure, and $V_{t,\mathrm{in}}(z,\omega)$ is the inward component of total particle velocity. It is an ideal means of characterizing materials within a frequency domain BEM as it relates the two surface unknowns, pressure and normal velocity, by a complex scalar allowing the problem to be simplified to having only one surface unknown.

The same relationship may be stated in the time domain as $p_t(0,t)=v_{t,\mathrm{in}}(0,t)^*z(t)$. However, a $z(t)$ found by inverse DFT of $Z(\omega)$ is typically noncompact in time and requires future values of $v_{t,\mathrm{in}}(0,t)$. This is due to the aggregation of cause and effect in the quantities $p_t(0,t)$ and $v_{t,\mathrm{in}}(0,t)$, and means that this form cannot be used with a time-marching solver. Further to this, the inverse Fourier transform of Eq. (6) appears to be a nontrivial operation.

Another approach is to relate the incoming and outgoing waves at the mouth of the well as $\Phi_{\mathrm{out}}(0,\omega)=\Phi_{\mathrm{in}}(0,\omega)W(\omega)$, where $W(\omega)=e^{i2kd}$ is the surface reflection coefficient. This too may be stated in the time domain:

$$\varphi_{\mathrm{out}}(0,t)=\varphi_{\mathrm{in}}(0,t)^*w(t), \tag{7}$$

where the time-invariant surface reflection kernel $w(t)$ is typically compact in time and expresses φ_{out} using only past values of φ_{in}, hence is suitable for use with a time-marching solver. For the well model above,

$$w(t)=\delta(t-2dc^{-1}). \tag{8}$$

Equations (7) and (8) will be used as the basis for the development of a time domain well mouth surface model.

A. Surface reflection boundary condition

Surface impedance is considered to vary spatially so $\hat{\mathbf{n}}_{\mathbf{x}}\cdot\nabla\Phi_t(\mathbf{x},\omega)=-i\omega\rho_0\Phi_t(\mathbf{x},\omega)Z(\mathbf{x},\omega)^{-1}$ holds for every point on S. Implicit in this is the assumption that the obstacle reacts locally. An equivalent time domain boundary condition may be written using the surface reflection kernel form of Eq. (7):

$$\varphi_{\mathrm{out}}(\mathbf{x},t)=\varphi_{\mathrm{in}}(\mathbf{x},t)^*w(\mathbf{x},t). \tag{9}$$

The total velocity potential used in Sec. II is the sum of the incoming and outgoing waves:

$$\varphi_t(\mathbf{x},t)=\varphi_{\mathrm{in}}(\mathbf{x},t)+\varphi_{\mathrm{out}}(\mathbf{x},t)=\varphi_{\mathrm{in}}(\mathbf{x},t)^*[\delta(t)+w(\mathbf{x},t)]. \tag{10}$$

This form suggests that $\varphi_{\mathrm{in}}(\mathbf{x},t)$ should be discretized as it is the fundamental surface unknown and $\varphi_{\mathrm{out}}(\mathbf{x},t)$ and $\varphi_t(\mathbf{x},t)$ are the secondary effects of its impingement on the obstacle. Similar statements may be found for total pressure and total normal velocity:

$$p_t(\mathbf{x},t)=-\rho_0\dot{\varphi}_t(\mathbf{x},t)=-\rho_0\frac{\partial}{\partial t}[\varphi_{\mathrm{in}}(\mathbf{x},t)^*[\delta(t)+w(\mathbf{x},t)]], \tag{11}$$

$$\begin{aligned}\hat{\mathbf{n}}_{\mathbf{x}}\cdot\nabla\varphi_t(\mathbf{x},t)&=\hat{\mathbf{n}}_{\mathbf{x}}\cdot\nabla[\varphi_{\mathrm{in}}(\mathbf{x},t)+\varphi_{\mathrm{out}}(\mathbf{x},t)]\\&=\frac{\partial}{\partial z}[\varphi_{\mathrm{in}}(\mathbf{x},t+zc^{-1})+\varphi_{\mathrm{out}}(\mathbf{x},t-zc^{-1})]\Big|_{z=0}\\&=\frac{1}{c}\frac{\partial}{\partial t}[\varphi_{\mathrm{in}}(\mathbf{x},t)^*[\delta(t)-w(\mathbf{x},t)]].\end{aligned} \tag{12}$$

The above statements have been written without specifying $w(\mathbf{x},t)$ in anticipation that they may be capable of describing a broader class of scattering obstacle. For the welled rigid surface model discussed $w(\mathbf{x},t)=\delta(t-2c^{-1}d(\mathbf{x}))$, where $d(\mathbf{x})=0$ for rigid nonwell surfaces sections. The sifting property of the delta function is exploited to simplify the above:

$$\varphi_t(\mathbf{x},t)=\varphi_{\mathrm{in}}(\mathbf{x},t)+\varphi_{\mathrm{in}}(\mathbf{x},t-2c^{-1}d(\mathbf{x})), \tag{13}$$

$$p_t(\mathbf{x},t)=-\rho_0[\dot{\varphi}_{\mathrm{in}}(\mathbf{x},t)+\dot{\varphi}_{\mathrm{in}}(\mathbf{x},t-2c^{-1}d(\mathbf{x}))], \tag{14}$$

$$\hat{\mathbf{n}}_{\mathbf{x}}\cdot\nabla\varphi_t(\mathbf{x},t)=\frac{1}{c}[\dot{\varphi}_{\mathrm{in}}(\mathbf{x},t)-\dot{\varphi}_{\mathrm{in}}(\mathbf{x},t-2c^{-1}d(\mathbf{x}))]. \tag{15}$$

These statements do not enforce that quantities are invariant over the cross section of the well as should be the case for plane waves; that restriction is left to the discretization scheme. Instead they state that the waves in the well are one dimensional, and that each point on the mouth of the well reacts locally. Radiation impedance of the well is accounted for in the boundary integral description of the problem. Substitution into Eq. (5) creates an operator that can calculate the sound scattered by an obstacle comprising rigid and welled sections while not supporting cavity resonances:

$$\begin{aligned}&(1-\alpha)\dot{\varphi}_i(\mathbf{x},t)+\alpha c\hat{\mathbf{n}}_{\mathbf{x}}\cdot\nabla\varphi_i(\mathbf{x},t)\\&\quad=L_c\{\varphi_{\mathrm{in}}(\mathbf{x},t)+\varphi_{\mathrm{in}}(\mathbf{x},t-2c^{-1}d(\mathbf{x}))\}\quad\text{if }\mathbf{x}\in S_-.\end{aligned} \tag{16}$$

In the following section, this new operator will be discretized to form a time-marching BEM.

IV. THE MARCHING-ON-IN-TIME METHOD

The surface quantities must be discretized in order for a solution to the boundary conditions on S to be found numerically. As suggested by Eq. (10), to ensure compatibility with

the new surface reflection boundary condition, $\varphi_{\text{in}}(\mathbf{x},t)$ will be discretized in preference to the usual $\varphi_t(\mathbf{x},t)$. Otherwise the discretization scheme follows that used in Ref. 12.

The surface S is partitioned into N_s flat elements denoted S_n, all small with respect to the anticipated spatial variation of the sound field, and time is discretized into N_t regular time-steps with duration Δ_t. Discretization of the incoming wave is achieved by approximating it by a weighted summation of basis functions:

$$\varphi_{\text{in}}(\mathbf{x},t) = \sum_{n=1}^{N_s} \sum_{i=1}^{N_t} w_{n,i} f_n(\mathbf{x}) T_i(t), \tag{17}$$

where $w_{n,i}$ are the discretization weights,

$$f_n(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in S_n \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

are the spatial basis functions, and

$$T_i(t) = T(t - i\Delta_t) \tag{19}$$

are the temporal basis functions, the latter being regularly delayed copies of the mother basis function $T(t)$. Currently $T(t)$ is chosen to be the piecewise polynomial used in Ref. 12.

This discretization scheme is substituted into Eq. (16) and the summations and weights are brought outside L_c. Collocation is performed in space and time to form a matrix equation; evaluation at $\mathbf{x}_m$ (the center of element S_m) and $t_j = j\Delta_t$ contributes a row to

$$\mathbf{Z}_0 \mathbf{w}_j = \mathbf{e}_j - \sum_{l=1}^{\infty} \mathbf{Z}_l \mathbf{w}_{j-l}, \tag{20}$$

where $l=j-i$ is the retardation index and the weights $\mathbf{w}_{i;n} = w_{n,i}$. The interaction matrices are defined as follows:

$$\mathbf{Z}_{l;m,n} = L_c\{f_n(\mathbf{x})[T_{j-l}(t_j) + T_{j-l}(t_j - 2c^{-1}d_n)]\}, \quad \mathbf{x} = \mathbf{x}_{m-}, \tag{21}$$

where d_n is the well depth of element S_n and equals zero for rigid nonwell elements. These are evaluated efficiently and accurately by regularization to contour integrals and adaptive numerical integration; details are included in the Appendix. The excitation vectors are evaluated as

$$\mathbf{e}_{j;m} = (1-\alpha)\dot{\varphi}_i(\mathbf{x}_m,t_j) + \alpha c \hat{\mathbf{n}}_{\mathbf{x}} \cdot \nabla \varphi_i(\mathbf{x}_m,t_j). \tag{22}$$

This algorithm is commonly referred to as the marching on in time (MOT) or "retarded potential" algorithm and intuitively possesses an iterative structure with sound traveling from element to element with a finite speed. It may more generally be considered to be a matrix solver between excitation coefficients and discretization weights, which exploits a pattern in the interaction matrices due to the regular temporal basis functions.

Discretization accuracy may be quantified spatially and temporally by considering the maximum frequency $\omega_{\max}$ present in the incident wave. The maximum phase variation over an element with largest dimension Δ_x in a time-step is $\omega_{\max}(\Delta_t + \Delta_x c^{-1})$. The logical assumption that spatial and temporal discretization error should be of similar magnitudes suggests the choice $\Delta_x \approx c\Delta_t$, as favored by Bluck and Walker.[8] This leads to nonzero off diagonals in the matrix $\mathbf{Z}_0$, necessitating a matrix solution at each time step. $\mathbf{Z}_0$ will in practice be very sparse and an iterative matrix solver seeded with the previous time-step's weights provides an efficient implementation.

In the following section, the new time domain well mouth surface model will be verified against frequency domain BEM implementations.

V. RESULTS

The new algorithm implements a time domain BEM model of a welled obstacle characterized by the new surface reflection boundary condition. This has required development of the integration scheme but the numerical machinery is otherwise as previously published. Verifying this algorithm's results will demonstrate that the new surface reflection boundary condition is performing correctly.

Verification is achieved by comparison with frequency domain BEM implementations which have previously been shown to accurately match experimental data.[36] Two scattering problems are considered, both of which involve obstacles possessing wells, the first being an object with uniform depth wells covering one face and the second being a Schroeder diffuser. A harmonic point source illuminates the surface for sufficient duration that the system reaches steady state and any instability has the opportunity to appear. The DFT is applied to the time domain data and the error versus the frequency domain BEM is quantified at the frequency of excitation. Using single frequency excitation is clearly an uninspiring application of the time domain BEM but is being done purely to achieve rigorous verification.

Two frequency domain BEM implementations are used. The closed body version models the well mouths as surface impedances according to Eq. (6). The open body version is capable of modeling the thin fins separating the wells so that the wells are modeled explicitly.

One mesh is used for each surface, and the time domain BEM is verified for a wide range of time-step durations defined by their relationship to spatial resolution, denoted implicitness $c\Delta_t\Delta_x^{-1}$. This is done because time-step duration has been associated with stability in many publications; evidently different values affect whether poles are perturbed into divergence. For each of these, a far-field harmonic point source, located 100 m distant normal to the obstacle, excites the system at a frequency such that the number of time-steps per excitation period $\beta = 2\pi(\Delta_t\omega)^{-1}$ assumes a range of predetermined values. For each combination, the error e between the time and frequency domain BEMs is calculated from the normalized mean complex difference between the respective source-to-collocation-point transfer functions at the excitation frequency:

$$e(\omega) = \frac{\Sigma_{m=1}^{N_s} |H_{\text{TD}}(\mathbf{x}_m,\omega) - H_{\text{FD}}(\mathbf{x}_m,\omega)|}{\Sigma_{m=1}^{N_s} |H_{\text{FD}}(\mathbf{x}_m,\omega)|}. \tag{23}$$

In the frequency domain, the transfer function H_{FD} is simply the total pressure divided by the source monopole pressure:

$$H_{\mathrm{FD}}(\mathbf{x},\omega)=\frac{P_t(\mathbf{x},\omega)}{P_{\mathrm{source}}}. \tag{24}$$

In the time domain H_{TD} is found by division of the DFT of the total velocity potential by the DFT of the source monopole potential:

$$H_{\mathrm{TD}}(\mathbf{x},\omega)=\frac{F\{\varphi_t(\mathbf{x},t)\}(\omega)}{F\{\varphi_{\mathrm{source}}(t)\}(\omega)}. \tag{25}$$

The first 50β (defined below) iterations are omitted from the DFT to allow the time domain solution to reach steady state. The next 100β iterations are chosen for DFT; this length maintains periodicity and eliminates windowing error. This error ratio is displayed for each integration type as a percentage contour plot between time-step implicitness and temporal resolution β.

A. Uniform welled body

This body is a box 1.0 m^2 by 0.5 m deep and its front face is covered by a lattice of wells all 0.1 m deep. Figure 4(a) shows the open surface mesh comprising 580 thin plate elements, and Fig. 4(b) shows the equivalent surface impedance mesh comprising 400 elements; the thin elements are partially transparent and the well mouth elements are colored darker. $\Delta_x=0.1$ m for both meshes. The surface impedance model requires roughly half the memory and computation time required by the thin plate model.

Figure 5 shows the error between the time domain and frequency domain models of the mesh in Fig. 4(b) versus time-step implicitness and temporal resolution. To the left of the figure where the time-step duration is explicit, spatial resolution is poor with respect to excitation wavelength so the accuracy of all BEM suffers. Toward the bottom of the figure, temporal resolution of the excitation frequency is poor; error here primarily originates from the time domain BEM. However, in the middle to upper right quadrant of the figure, discretization error is low and good agreement occurs. No instability is witnessed for any time-step duration indicating that the CFIE operator has successfully avoided supporting any cavity resonances even when combined with the new well mouth surface model.

Figure 6 shows the interference patterns that occur between incident and scattered sound as further evidence that the well mouth surface model is behaving as expected. The receivers are arranged in a vertical line that starts behind the obstacle (left of the figure), passes through its center, and emerges at the front (right of the figure). They are spaced such that none touches the surface. The magnitude of the source to receiver transfer function H is plotted in decibel versus the receiver z coordinate. Data series are shown for the surface impedance mesh [Fig. 4(b)] modeled by the time domain BEM (TD impedance) and the frequency domain BEM for closed surfaces (FD impedance), and for the thin plate mesh [Fig. 4(a)] modeled by the frequency domain BEM for open surfaces (FD thin plate). The vertical lines at $z=0.0$ and $z=-0.5$ indicate the front and back of the obstacle and the shaded area indicates the wells of the mixed mesh. Time-step duration was chosen such that $\Delta_x=c\Delta_t$.

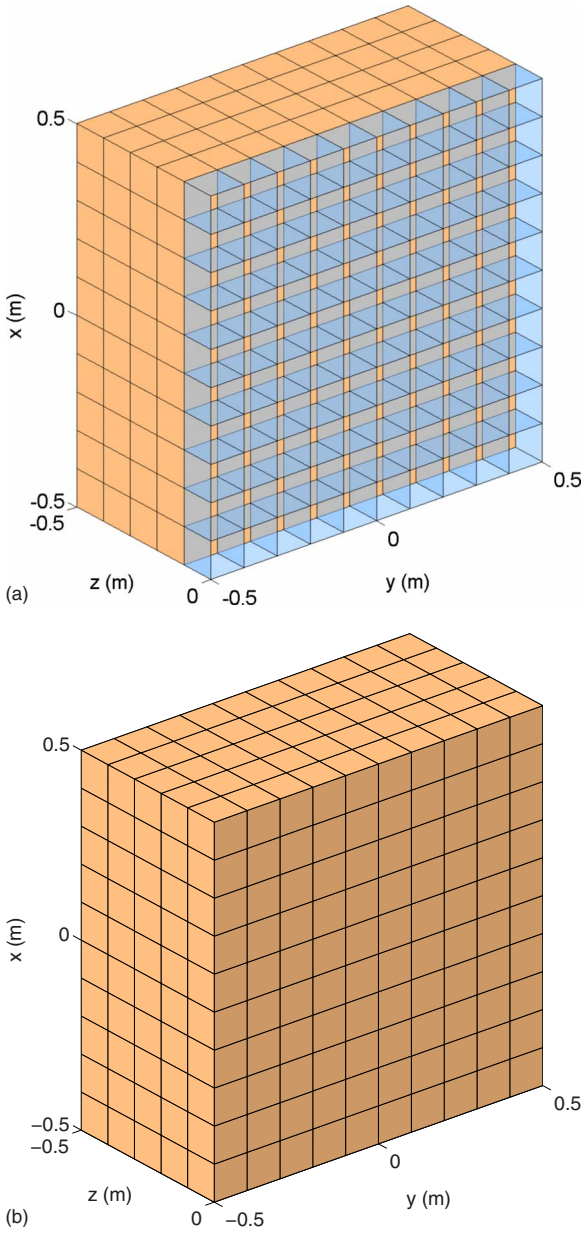


FIG. 4. (Color online) Meshes of a body with uniform depth wells on its front face: (a) thin plate model and (b) surface impedance model.

Interference effects between the incident and scattered waves are evident in front of the surface and in this region there is excellent agreement between the time and frequency domain algorithms. The BEM for open surfaces is seen to extend the interference patterns into the welled region and its surface normal gradient approaches zero as expected from the rigid well floors. Inside the surface, the frequency domain BEM for closed surfaces achieves the best cancellation, but the results still confirm that the time domain surface reflection response boundary condition does not permit sound to flow into the cavity. In the shadow region behind the obstacle, all models, roughly agree but there is no apparent interference behavior.

This modeling problem has shown excellent stability and agreement with a verified frequency domain BEM on a simple surface with welled sections. A more complex real-world device will now be modeled.

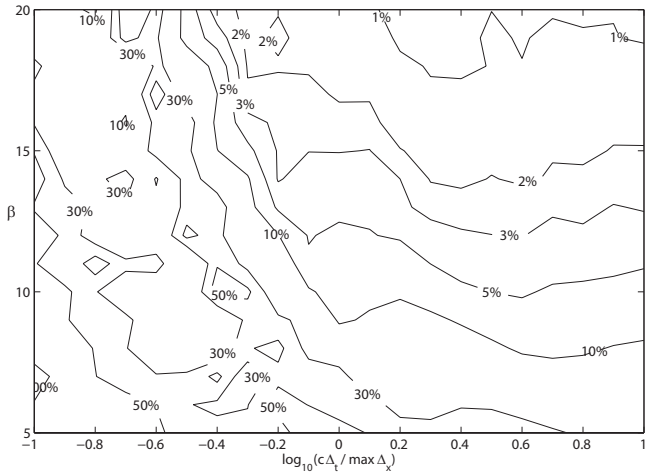


FIG. 5. Error between time and frequency domain BEMs for closed surfaces both modeling the uniform welled surface.

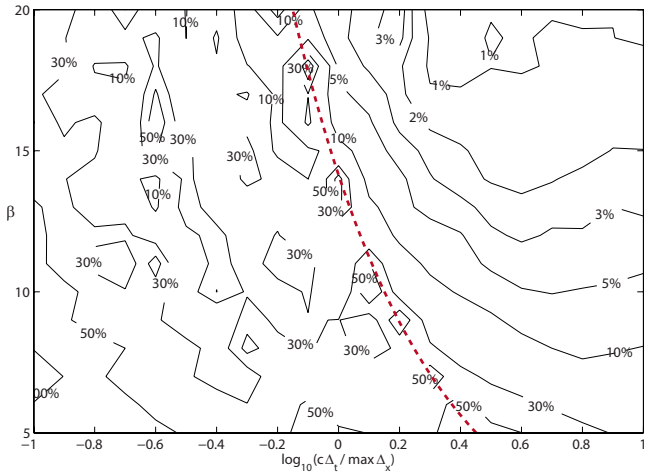


FIG. 7. (Color online) Error between the time and frequency domain BEM for closed surfaces on the surface impedance model of the QRD. The dashed line indicates the first modal frequency along the wells of the QRD.

B. Quadratic residue diffuser

This device modeled has a design wavelength of 1.4 m, a well width of 0.3 m, a height of 1.0 m and follows the quadratic residue depth sequence [2 4 1 0 1 4 2]. A diagram of the real device and the equivalent surface impedance model were shown in Figs. 1(a) and 1(b), respectively. $\Delta_x = 0.1$ m for both meshes giving them 972 and 792 elements, respectively; consequentially, the surface impedance model requires roughly two-thirds of the memory and computation time of the thin plate model.

The relationship between the quadratic residue diffuser (QRD) meshes does not exactly mirror that between the uniform welled body meshes. Specifically, the well mouths of the surface impedance model mesh have been discretized into elements with the same spatial resolution as the rigid parts of the surface; this is a standard technique used to mesh compliant surfaces in the frequency domain. Each element reacts locally which is equivalent to having the wells partitioned by a lattice structure similar to that present on the front face of Fig. 4. This is not quite equivalent to either the real device, the model in Fig. 1(a), or Schroeder's plane wave model, but it has been shown to produce reasonable results.[3]

Figure 7 shows the error between the time and frequency domain implementations of the surface impedance model of the QRD depicted in Fig. 1(b). It is plotted versus time-step implicitness and temporal resolution. The trends are the same as observed for the uniform welled surface, with universal stability and good agreement in the upper right quadrant, indicating that the new algorithm is functioning correctly. The first modal frequency along the wells of the QRD (171.5 Hz, $\lambda = 2$ m) is indicated by a dashed line and increased error can be observed close to it. This was an unexpected result since the well mouth impedance model effectively partitions the well, and its significance is an ongoing research question.

Figure 8 shows the magnitude of the source to receiver scattered sound transfer function in decibel versus receiver angle relative to the surface normal. This is calculated according to Eqs. (24) and (25) with the modification that scattered pressure and velocity potential are used in place of their total sound counterparts. The 91 receivers are uniformly spaced in a 5 m radius arc located in the primary scattering plane of the QRD. Data series are shown for the surface impedance mesh [Fig. 1(b)] modeled by the time domain

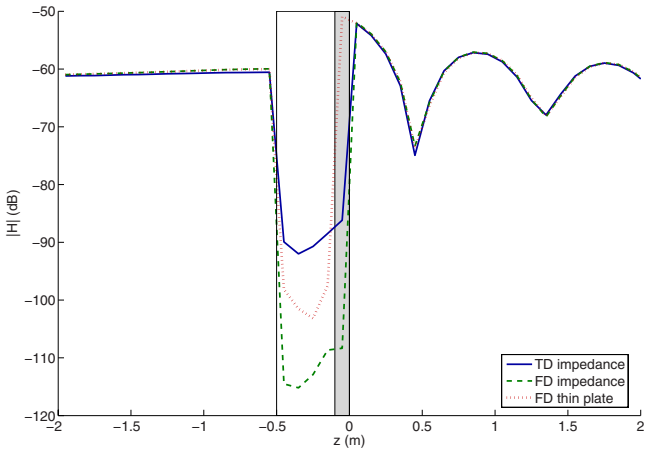


FIG. 6. (Color online) Total receiver sound though the uniform welled surface at 202 Hz $\equiv \beta = 12$. The vertical lines indicate the front and back of the obstacle and shading the wells on its front face.

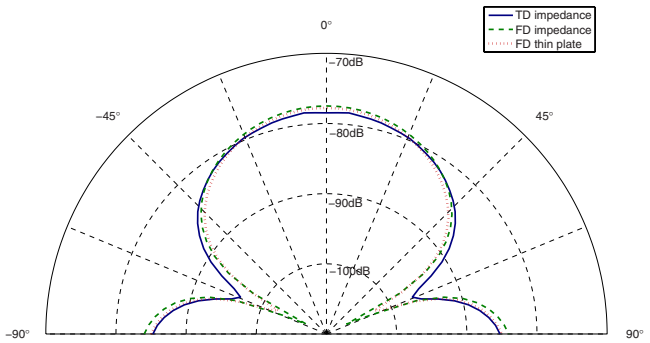


FIG. 8. (Color online) Scattered sound (dB) at a receiver arc 5 m from the QRD at 202 Hz $\equiv \beta = 12$.

BEM (TD impedance) and the frequency domain BEM for closed surfaces (FD impedance), and for the thin plate mesh [Fig. 1(a)] modeled by the frequency domain BEM for open surfaces (FD thin plate). Time-step duration and excitation frequency were chosen such that $\Delta_x = c\Delta_t$ and $\beta = 12$.

Agreement between the models is good, though it is recognized that the figure has a broad range of scale and unlike Fig. 7 does not consider phase agreements. Despite the fact that this frequency is slightly below the design frequency of the diffuser (245 Hz) some grating lobe behavior is visible and predicted by all models. Cancellation of incident and scattered waves at the scattering nulls ($\approx \pm 70°$) is more complete for the frequency domain results, mirroring the trend inside the body of the uniform welled surface (Fig. 6), suggesting that the frequency domain algorithms are more accurate for time-harmonic problems.

This modeling problem has shown excellent stability and agreement with verified frequency domain BEMs for a geometrically complex real-world acoustic device. In the application of welled surfaces, the surface reflection boundary condition is achieving in the time domain what surface impedance achieves in the frequency domain.

VI. CONCLUSIONS

The investigation sought to transfer the technique of modeling the well mouths of a Schroeder diffuser as compliant surfaces from the frequency to the time domain. A direct inverse Fourier transform of surface impedance is unsuitable as its convolution kernels are typically noncompact in time and requires future data not available to a time-marching solver. Instead the response of the well to an incoming wave was described by its convolution with a time-invariant kernel denoted surface reflection response, which for a well is extremely compact in time and only requires past data. Pressure and normal velocity at the mouth of the well are readily found from the incoming velocity potential so this was discretized in preference to the more usual choice of total velocity potential. The new surface model was implemented and the algorithm verified on two welled surfaces, one of which was a Schroeder diffuser. Agreement with previously verified frequency domain BEM implementations was good.

It is hoped that with further research this surface model may be generalized to include other classes of compliant obstacle currently modeled as locally reacting surface impedances in the frequency domain. One obstacle is the absence of data for the surface reflection response model when the obstacle is too complex to be considered analytically. Third octave averaged absorption coefficients are usually measured and quoted for most materials, and do not contain enough data to directly reconstruct the surface reflection kernel; instead some form of estimation is required. Drumm and Lam[31] and Fung *et al.*[32] have both suggested approaches to replacing the convolutions appearing in FDTD surface models with simple filters fitted to the absorption data, and their approaches may also be effective for time domain BEM. Further potential lies in the application of modeling materials with nonlinear and time-variant properties, for which time-harmonic models do not exist.

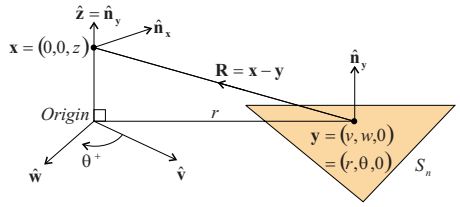


FIG. 9. (Color online) Contour integration geometry and coordinate systems.

ACKNOWLEDGMENTS

This project was been funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant No. GR/P01144/01.

APPENDIX: NUMERICAL EVALUATION OF INTERACTION COEFFICIENTS

Accurate evaluation of the interaction coefficients defined in Eq. (21) is fundamental to the accuracy and stability of the algorithm. The temporal basis function chosen has discontinuous derivatives which cause discontinuities and delta functions in the surface integrands, making them unsuitable for solution by Gaussian integration. In addition, the integrand is singular so element self-interaction need often be considered as a special case.

The implementation herein exploits the flat elements and piecewise-constant spatial basis functions to permit regularization of all integrands by coordinate transformation,[25,37] such that the collocation point is no longer a special case. The radial component of integration is performed analytically, leaving the remaining numerical integration a one-dimensional contour integral. This allows an adaptive numerical integration scheme to be used, specifically Simpson integration with Romberg extrapolation. An absolute termination criterion was used, meaning that larger more significant interaction coefficients were evaluated with higher precision than smaller less significant ones. This process is arbitrarily accurate, has better computational cost scaling than two-dimensional integration, and allows the same integration routine to be used for all element pairs as effort is automatically concentrated where necessary.

In order to clarify the conversion of the surface integral over S_n into nested integrals two new coordinate systems will be used; one is a Cartesian system (v,w,z) and one a cylindrical polar system (r,θ,z), both shown in Fig. 9. The origin and positive z direction are the same in both coordinate systems. The origin is defined as the projection of the collocation point $\mathbf{x}$ into the plane of S_n and the positive z direction is specified by $\hat{\mathbf{n}}_\mathbf{y}$. The positive v direction is defined as the projection of $\hat{\mathbf{n}}_\mathbf{x}$ into the plane of S_n, such that $\hat{\mathbf{w}} \cdot \hat{\mathbf{n}}_\mathbf{x} = 0$. The positive theta direction is defined such that $v = r\cos(\theta)$ and $w = r\sin(\theta)$ in the conventional way. The variable z refers for the z coordinate of the collocation point $\mathbf{x}$ and any reference to v, w, r, or θ implies the integration point $\mathbf{y}$.

The interaction coefficients are now evaluated according to the following expression, being the sum of integrals over the edges of S_n and a contribution from the origin:

$$\mathbf{Z}_{l;m,n} = \frac{1}{4\pi}\sum_{\text{edges}}\int_0^1 \left\{\left[\left(1-\alpha-\alpha\hat{\mathbf{n}}_{\mathbf{x}}\cdot\hat{\mathbf{n}}_{\mathbf{y}}\frac{z}{R}\right)\frac{d\theta}{d\mu}+\alpha\hat{\mathbf{n}}_{\mathbf{x}}\cdot\hat{\mathbf{v}}\frac{1}{R}\frac{dw}{d\mu}\right]\left[\left(\frac{z}{R}-1\right)\dot{T}_{j-l}\left(t_j-\frac{R}{c}\right)+\left(\frac{z}{R}+1\right)\dot{T}_{j-l}\left(t_j-\frac{R+2d_n}{c}\right)\right]\right.$$
$$\left.+\alpha\frac{c}{R}\left[\hat{\mathbf{n}}_{\mathbf{x}}\cdot\hat{\mathbf{n}}_{\mathbf{y}}\left(1-\frac{z^2}{R^2}\right)\frac{d\theta}{d\mu}+\hat{\mathbf{n}}_{\mathbf{x}}\cdot\hat{\mathbf{v}}\frac{z}{R^2}\frac{dw}{d\mu}\right]\left[T_{j-l}\left(t_j-\frac{R}{c}\right)+T_{j-l}\left(t_j-\frac{R+2d_n}{c}\right)\right]\right\}d\mu$$
$$-\frac{\theta_{\text{origin}}}{4\pi}\left[1-\alpha-\alpha\hat{\mathbf{n}}_{\mathbf{x}}\cdot\hat{\mathbf{n}}_{\mathbf{y}}\frac{z}{|z|}\right]\left[\left(\frac{z}{|z|}-1\right)\dot{T}_{j-l}\left(t_j-\frac{|z|}{c}\right)+\left(\frac{z}{|z|}+1\right)\dot{T}_{j-l}\left(t_j-\frac{|z|+2d_n}{c}\right)\right] \tag{A1}$$

Numerical integration is with respect to μ, a dimensionless edge position coefficient varying from zero at the start vertex to one at the end vertex. For an edge **e** the partial differentials between this and the geometric integration variables are found as follows, where $r_\perp$ is the minimum (perpendicular) distance from the origin to the line of edge **e**:

$$\frac{d\theta}{d\mu} = \frac{|\mathbf{e}|r_\perp}{r^2}\text{sign}(\hat{\mathbf{e}}\cdot\hat{\boldsymbol{\theta}}), \tag{A2}$$

$$\frac{dw}{d\mu} = \hat{\mathbf{w}}\cdot\mathbf{e}, \tag{A3}$$

θ_{origin} is the angle the edges of S_n make around the origin. This is zero if the origin is outside S_n and 2π if S_n contains the origin. If the origin lies on an edge, θ_{origin} will equal the enclosed angle, intersection of one edge implies $\theta_{\text{origin}}=\pi$, and intersection of a corner implies θ_{origin} will equal the acute angle between the adjoining edges.

[1]P. D'Antonio and T. J. Cox, "Diffusor application in rooms," Appl. Acoust. **60**, 113–142 (2000).
[2]Audio Engineering Society Inc., "AES information document for room acoustics and sound reinforcement systems—Characterisation and measurement of surface scattering uniformity," J. Audio Eng. Soc. AES-4id-2001 (r2007).
[3]T. J. Cox and Y. W. Lam, "Prediction and evaluation of the scattering from quadratic residue diffusers," J. Acoust. Soc. Am. **95**, 297–305 (1994).
[4]T. J. Cox and P. D'Antonio, *Acoustic Absorbers and Diffusers* (Spon, London, 2004).
[5]M. B. Friedman and R. P. Shaw, "Diffraction of pulses by cylindrical obstacles of arbitrary cross section," ASME J. Appl. Mech. **29**, 40–46 (1962).
[6]B. P. Rynne, "Instabilities in time marching methods for scattering problems," Electromagnetics **6**, 129–144 (1986).
[7]B. P. Rynne and P. D. Smith, "Stability of time marching algorithms for the electric field integral equation," Electromagn. Waves **4**, 1181–1205 (1990).
[8]S. J. Dodson, S. P. Walker, and M. J. Bluck, "Implicitness and stability of time domain integral equation scattering analysis," ACES. J. **13**, 291–301 (1998).
[9]P. D. Smith, "Instabilities in time marching methods for scattering: cause and rectification," Electromagnetics **10**, 439–451 (1990).
[10]H. Wang, D. J. Henwood, P. J. Harris, and R. Chakrabarti, "Concerning the cause of instability in time-stepping boundary element methods applied to the exterior acoustic problem," J. Sound Vib. **305**, 289–297 (2007).
[11]A. J. Burton and G. F. Miller, "The application of integral equation methods to the numerical solution of some exterior boundary-value problems," Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).
[12]A. A. Ergin, B. Shanker, and E. Michielssen, "Analysis of transient wave scattering from rigid bodies using a Burton–Miller approach," J. Acoust. Soc. Am. **106**, 2396–2404 (1999).
[13]A. A. Ergin, B. Shanker, and E. Michielssen, "Fast analysis of transient acoustic wave scattering from rigid bodies using the multilevel plane wave time domain algorithm," J. Acoust. Soc. Am. **107**, 1168–1178 (2000).
[14]A. E. Yilmaz, J.-M. Jina, and E. Michielssen, "A fast Fourier transform accelerated marching-on-in-time algorithm for electromagnetic analysis," Electromagnetics **21**, 181–197 (2001).
[15]G. C. Herman and P. M. van den Berg, "A least squares iteratively technique for solving time-domain scattering problems," J. Acoust. Soc. Am. **72**, 1947–1953 (1982).
[16]Y. Shifman and Y. Leviatan, "On the use of spatio-temporal multiresolution analysis in method of moments solutions of transient electromagnetic scattering," IEEE Trans. Antennas Propag. **49**, 1123–1129 (2001).
[17]Y. K. Chung, T. K. Sarkar, B. H. Jung, M. Salazar-Palma, Z. Ji, S. Jang, and K. Kim, "Solution of time domain electric field integral equations using the Laguerre polynomials," IEEE Trans. Antennas Propag. **52**, 2319–2328 (2004).
[18]Z. Ji, T. K. Sarkar, B. H. Jung, M. Yuan, and M. Salazar-Palma, "Solving time domain electric field integral equation without the time variable," IEEE Trans. Antennas Propag. **54**, 258–262 (2006).
[19]S. Barmada, "Improving the performance of the boundary element method with time-dependent fundamental solutions by the use of a wavelet expansion in the time domain," Int. J. Numer. Methods Eng. **71**, 363–378 (2007).
[20]M. R. Schroeder, "Diffuse sound reflection by maximum-length sequences," J. Acoust. Soc. Am. **57**, 149–150 (1975).
[21]M. R. Schroeder, "Binaural dissimilarity and optimum ceilings for concert halls: More lateral sound diffusion," J. Acoust. Soc. Am. **65**, 958–963 (1979).
[22]A. Farina, "A new method for measuring the scattering coefficient and the diffusion coefficient of panels," **86**, 928–942 (2000).
[23]J. Redondo, R. Picó, B. Roig, and M. R. Avis, "Time domain simulation of sound diffusers using finite-difference schemes," **93**, 611–622 (2007).
[24]R. Martinez, "The thin-shape breakdown (TSB) of the Helmholtz integral equation," J. Acoust. Soc. Am. **90**, 2728–2738 (1991).
[25]Y. Kawai and T. Terai, "A numerical method for the calculation of transient acoustic scattering from thin rigid plates," J. Sound Vib. **141**, 83–96 (1990).
[26]T. J. Cox and Y. W. Lam, "Prediction and evaluation of the scattering from quadratic residue diffusers," J. Acoust. Soc. Am. **95**, 297–305 (1994).
[27]T. Ha-Duong, B. Ludwig, and I. Terrasse, "A Galerkin BEM for transient acoustic scattering by an absorbing obstacle," Int. J. Numer. Methods Eng. **57**, 1845–1882 (2003).
[28]P. H. L. Gronenboom, "The applications of boundary elements to steady and unsteady potential fluid flow problems in two and three dimensions," Appl. Math. Model. **6**, 35–40 (1982).
[29]R. P. Shaw, "Diffraction of acoustic pulses by obstacles of arbitrary shape with a robin boundary condition," J. Acoust. Soc. Am. **41**, 855–859 (1967).
[30]C. K. W. Tam and L. Auriault, "Time-domain impedance boundary conditions for computational aeroacoustics," AIAA J. **34**, 917–923 (1996).
[31]I. A. Drumm and Y. W. Lam, "Development and assessment of a finite difference time domain room acoustic prediction model that uses hall data in popular formats," Proceedings of Internoise, Turkey, 2007.
[32]K.-Y. Fung, H. Ju, and B.-P. Tallapragada, "Impedance and its time-domain extensions," AIAA J. **38**, 30–38 (2000).
[33]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Application* (McGraw-Hill, London, 1981).
[34]D. A. Vechinski and S. M. Rao, "Transient scattering from dielectric cylinders: E-field, h-field and combined-field solutions," Radio Sci. **27**, 611–622 (1992).

[35]D. J. Chappell, P. J. Harris, D. Henwood, and R. Chakrabarti, "A stable boundary integral equation method for modeling transient acoustic radiation," J. Acoust. Soc. Am. **120**, 74–80 (2006).
[36]T. J. Cox and Y. W. Lam, "Evaluation of methods for predicting the scattering from simple rigid panels," Appl. Acoust. **40**, 123–140 (1993).
[37]Y. Ding, A. Forestier, and T. Ha-Duong, "A Galerkin scheme for the time domain integral equation of acoustic scattering from a hard surface," J. Acoust. Soc. Am. **86**, 1566–1572 (1989).

Analytical prediction of the breakout noise from a rectangular cavity with one compliant wall

B. Venkatesham,[a] Mayank Tiwari,[a] and M. L. Munjal[b]
Facility for Research in Technical Acoustics, Department of Mechanical Engineering, Indian Institute of Science, Bangalore 560012, India

(Received 12 December 2007; revised 4 June 2008; accepted 28 July 2008)

This paper describes an analytical calculation of breakout noise by incorporating three-dimensional effects along with the acoustical and structural wave coupling phenomena. The breakout noise phenomena from cavities are important at low frequencies, and the coupling between acoustic waves and structural waves plays a critical role in prediction of the transverse transmission loss. The first step in the breakout noise prediction is to calculate the inside cavity pressure field and the normal cavity wall vibration by using an impedance-mobility approach, which results in a compact matrix formulation. The second step is to calculate the radiated sound power from an unbaffled plate formulation that poses formidable challenges on computational time. The proposed formulation helps in reducing the computational time substantially by converting quadruple integrals into single integrals using an appropriate coordinate transformation technique. Analytical results are validated with the finite element/boundary element numerical models.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977671]

PACS number(s): 43.55.Rg, 43.20.Ks, 43.20.Rz, 43.20.Tb [LLT] Pages: 2952–2962

I. INTRODUCTION

Interaction between acoustic space and flexible structure is an important problem in the field of acoustics for different practical applications such as heating, ventilation and air-conditioning (HVAC) duct breakout noise reduction, thin wall enclosures, aircraft skin panel vibration analysis, active noise control from panels, sound transmission measurement in building acoustics, etc. A literature survey of the structural-acoustic coupling shows that a large amount of work has been done in calculating free and forced responses of different systems. Dowell and Voss[1] were the first investigators of modal response of a cavity-backed plate system. Pretlove[2–4] analyzed the free and forced responses of a cavity-plate system and observed that cavity plays an important role in the plate response. Further investigations by Pan and co-workers,[5–7] Guy,[8] Guy and Bhattacharya,[9] and Qaisi,[10] among others, progressively improved theoretical knowledge and physical insights of the problem. Lyon[11] studied the noise reduction produced by a small rigid enclosure with one flexible wall at different frequencies and explained cavity behavior. Dowell *et al.*[12] developed a comprehensive theoretical model for coupled responses in a structural-acoustic coupled system.

Kim and Brennan[13] re-examined the formulation used by Dowell *et al.*[12] and expressed the same in the impedance and mobility terms. This method provides coupled response in terms of the modal characteristics of the uncoupled structural and acoustical systems, and the equations describing the complete system are expressed in terms of matrices. It is easy to do computations with matrices and implement the same on a computer. The off-diagonal terms in the coupling matrices are due to the coupled acoustic impedance and coupled structural mobility. This method gives the inside pressure and compliant wall vibration displacement simultaneously. This can be used in the Kirchhoff–Helmholtz (KH) integral to predict the radiated power from the cavity-backed plate. The radiation impedance term involves a quadruple integral. Solving this integral is a quite complex and computationally expensive exercise. Li and Gilbeling[14] presented a two-step method to solve these quadruple integrals for a rectangular baffled plate (BP). The first step is to convert the quadruple integral into double integrals by Cartesian coordinate transformation and the second step is to convert double integrals into single integrals by polar coordinate transformation. Nelisse *et al.*[15,16] discussed the sound radiation of a BP and/or an unbaffled plate (UBP) with arbitrary boundary conditions. They proposed a generic algorithm involving the twofold integrals for mass and stiffness matrix coefficients and solved these integrals numerically.

It may be noted from the foregoing literature survey that there has hardly been any serious attempt at prediction of the breakout noise from a rectangular cavity. This lacuna is proposed to be filled up in the present paper. To start with, a rectangular cavity with one compliant wall is investigated. An analytical solution is provided for rectangular cavity with one of the walls being compliant or flexible; the remaining sides or walls are assumed to be rigid. The first step in this method is to calculate the inside acoustic pressure field of the cavity and compliant wall vibration using an impedance-mobility compact matrix (IMCM) approach, which results in a compact matrix formulation. The second step is to predict the noise radiation from a compliant wall using the KH integral. Finally, one can predict the transverse transmission loss using the incidence power inside the cavity and the radiated power values. Another objective of the current inves-

[a]Also at GE Global Research, Bangalore, India.
[b]Author to whom correspondence should be addressed. Electronic mail: munjal@mecheng.iisc.ernet.in

 © 2008 Acoustical Society of America

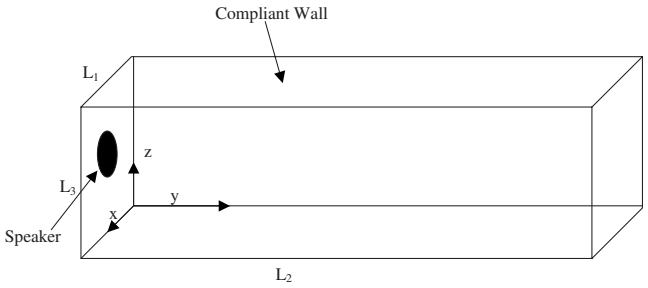


FIG. 1. Rectangular cavity with one complaint wall, excited by a speaker.

tigation is to understand the coupling phenomenon and its effect on the compliant wall vibration displacement. The compliant wall vibration velocity is a linear combination of the *in vacuo* flexural modes of the plate. It is substituted into the KH integral to predict the sound pressure radiated by the vibrating compliant wall without considering the cavity volume effect. The radiated sound power can be obtained by integrating the acoustic intensity over the surface of the flexible wall and it gives the sound power in terms of radiation impedance. In the current analysis, the self-radiation impedance and mutual radiation impedance terms are incorporated to consider the mode coupling effect on sound power radiation. The quadruple integrals involved in the mass and stiffness matrices are converted into single integrals by using a two-step coordinate transformation. It is shown to be much faster than solving the twofold integrals numerically.

II. CALCULATION OF THE INSIDE CAVITY PRESSURE AND COMPLIANT WALL VIBRATION

Figure 1 shows the schematic of a rectangular cavity with one flexible wall; all other walls or sides are assumed to be rigid in this investigation. One side of the cavity is attached with an acoustic source. In the IMCM approach, it is presumed that coupled response can be described in terms of finite sets of the uncoupled acoustic cavity and structural modes.[13] A strong coupling is assumed between the cavity and the flexible plate, and a weak coupling between the flexible plate and the outside radiated acoustic field.

General equations of the inside cavity pressure p at location $\mathbf{x}$ and the compliant wall vibration velocity w at location $\mathbf{z}$ on the flexible plate for the uncoupled cavity modes N and structural modes M are

$$p(\mathbf{x},\omega) = \sum_{n=1}^{N} \psi_n(\mathbf{x}) a_n(\omega) = \boldsymbol{\psi}^T \mathbf{a}, \tag{1a}$$

$$w(\mathbf{z},\omega) = \sum_{m=1}^{M} \phi_m(\mathbf{z}) b_m(\omega) = \boldsymbol{\varphi}^T \mathbf{b}, \tag{1b}$$

where $\psi_n(\mathbf{x})$ is the uncoupled acoustic mode shape function, $a_n(\omega)$ is the complex amplitude of the nth acoustic pressure mode, $\phi_m(\mathbf{z})$ is the uncoupled vibration mode shape function, and $b_m(\omega)$ is the complex amplitude of the mth vibration velocity mode. The nth acoustic mode shape function for a rigid-walled rectangular cavity normalized by its volume is given by[17]

$$\psi_n(\mathbf{x}) = \sqrt{e_1 e_2 e_3}\cos\left(\frac{n_1\pi x}{L_1}\right)\cos\left(\frac{n_2\pi y}{L_2}\right)\cos\left(\frac{n_3\pi z}{L_3}\right), \tag{2}$$

where n denotes the trio of modal integers n_1, n_2, and n_3. $n=1$ means $n_1=n_2=n_3=0$, the mode referred as the (0,0,0) mode that corresponds to the purely compliant behavior of the air in the enclosure due to its bulk modulus. L_1, L_2, and L_3 are dimensions of the rectangular cavity in the x, y, and z coordinate directions, as shown in Fig. 1. The normalization factors are given by $e_i=1$ if $n_i=0$ and $e_i=2$ if $n_i \geq 1$, where the subscript i can be 1, 2, and 3.

The structural mode shape function for a simply supported rectangular plate normalized by its surface area is[13]

$$\phi_m(\mathbf{z}) = 2\sin\left(\frac{m_1\pi x}{L_1}\right)\sin\left(\frac{m_2\pi y}{L_2}\right), \tag{3}$$

where m_1 and m_2 are mode numbers with positive integers.

The modal acoustic pressure vector $\mathbf{a}$ can be expressed in the matrix form as

$$\mathbf{a} = \mathbf{Z_a}(\mathbf{q} + \mathbf{q_s}), \tag{4}$$

where $\mathbf{Z_a}=\mathbf{A}\rho_0 c_0^2/V$ is an $(N\times N)$ diagonal matrix defined as the uncoupled acoustic modal impedance matrix: $\mathbf{q_s}=\mathbf{Cb}$; $\mathbf{q}$ is the N-length modal source strength vector and $\mathbf{q_s}$ is the modal source vector due to vibration of the structure. Matrix $\mathbf{A}$ is an $(N\times N)$ diagonal matrix. It contains acoustic mode resonance terms.[13] $\mathbf{C}$ is an $(N\times M)$ matrix defined as the coupling coefficient between the nth acoustic mode and the mth structural mode. ρ_0 and c_0 denote the mass density and the speed of sound in air, respectively.

The modal vibration amplitude vector $\mathbf{b}$ can be expressed in the matrix form as[13]

$$\mathbf{b} = \mathbf{Y_s}(\mathbf{g} - \mathbf{g_a}), \tag{5a}$$

$$\mathbf{g_a} = \mathbf{C^T a}, \tag{5b}$$

where $\mathbf{g}$ is the generalized modal force vector due to the external force distribution, $\mathbf{g_a}$ is the modal force vector acting on the acoustic system, and $\mathbf{Y_s}=B/(\rho_s h S_f)$ is the $(M\times M)$ diagonal matrix defined as the uncoupled structural modal mobility matrix. B is an $(M\times M)$ diagonal matrix consisting of a structural mode resonance term.[18] ρ_s and h denote the density of the plate material and the thickness of the plate, respectively.

Combining Eqs. (4) and (5a) yields the acoustic and structural modal amplitude vectors $\mathbf{a}$ and $\mathbf{b}$ in terms of the modal excitation vectors $\mathbf{q}$ and $\mathbf{g}$ as follows:

$$\mathbf{a} = (\mathbf{I} + \mathbf{Z_a C Y_s C^T})^{-1}\mathbf{Z_a}(\mathbf{q} + \mathbf{C Y_s g}), \tag{6a}$$

$$\mathbf{b} = (\mathbf{I} + \mathbf{Y_s C^T Z_a C})^{-1}\mathbf{Y_s}(\mathbf{g} - \mathbf{C^T Z_a q}). \tag{6b}$$

The coupled natural frequencies of the cavity-panel system are calculated by rearranging the expressions of complex amplitude of the nth acoustic mode and the vibration amplitude of the mth mode.[12,19] Considering N uncoupled acoustic modes and M uncoupled structural modes gives linear algebraic equations. These equations are arranged in the matrix form, and then application of zero excitation gives free vibration analysis of coupled system in the final form

$$(\lambda^2\mathbf{M} + \lambda\mathbf{L} + \mathbf{S})\mathbf{X} = 0, \tag{7}$$

where $\mathbf{M}$ is the inertia matrix, $\mathbf{L}$ is the coupling matrix, $\mathbf{S}$ is the stiffness matrix, $\lambda = j\omega$ is the eigenvalue, and $\mathbf{X} = [\mathbf{a}\ \ \mathbf{b}]$.[20]

By means of a matrix transformation, the nonstandard eigenequation (7) can be changed into the standard real eigenequation, by introducing $\mathbf{Y} = \lambda\mathbf{X}$. Then Eq. (7) reduces to the following standard eigenequation, which can be solved by using any commercial mathematical package:

$$\begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{S} & -\mathbf{M}^{-1}\mathbf{L} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}, \tag{8}$$

where

$$\mathbf{M} = \begin{bmatrix} \frac{V}{\rho_0 c_0^2}[\mathbf{I}]_{N\times N} & \mathbf{0} \\ \mathbf{0} & \rho_s h s_f [\mathbf{I}]_{M\times M} \end{bmatrix}, \tag{9}$$

$$\mathbf{S} = \begin{bmatrix} \mathbf{S_{11}} & \mathbf{0} \\ \mathbf{0} & \mathbf{S_{22}} \end{bmatrix}, \tag{10}$$

$$\mathbf{L} = \begin{bmatrix} \mathbf{L_{11}} & -\mathbf{C} \\ \mathbf{C^T} & \mathbf{L_{22}} \end{bmatrix}, \tag{11}$$

$$\mathbf{S_{11}} = \frac{V}{\rho_0 c_0^2} \begin{bmatrix} \omega_1^2 & \cdots & \mathbf{0} \\ | & \omega_n^2 & | \\ \mathbf{0} & \cdots & \omega_N^2 \end{bmatrix}_{N\times N},$$

$$\mathbf{S_{22}} = \rho_s h s_f \begin{bmatrix} \omega_1^2 & \cdots & \mathbf{0} \\ | & \omega_m^2 & | \\ \mathbf{0} & \cdots & \omega_{\mathbf{M}}^2 \end{bmatrix}_{M\times M},$$

$$\mathbf{L_{11}} = \frac{2V}{\rho_0 c_0^2} \begin{bmatrix} \zeta_1\omega_1 & \cdots & \mathbf{0} \\ | & \zeta_n\omega_n & | \\ \mathbf{0} & \cdots & \zeta_N\omega_N \end{bmatrix}_{N\times N},$$

$$\mathbf{L_{22}} = 2\rho_s h s_f \begin{bmatrix} \zeta_1\omega_1 & \cdots & \mathbf{0} \\ | & \zeta_m\omega_m & | \\ \mathbf{0} & \cdots & \zeta_M\omega_M \end{bmatrix}_{M\times M}.$$

ω_N and ζ_N are the natural frequency and damping ratio of the Nth acoustic mode, respectively.

III. CALCULATION OF THE ACOUSTIC POWER RADIATION

Calculation of acoustic power radiation from the cavity-panel system has been done with the assumption of cavity volume not contributing to the noise radiation from the panel. Generally, the cavity volume adds stiffness to the panel in the lower frequency region. This is verified by using numerical calculations of the cavity-panel system. Neglecting the cavity volume effect at low frequencies, the panel can be modeled as an UBP with the applied velocity excitation. This assumption, however, has a limitation in the lower frequency region, as is shown later.

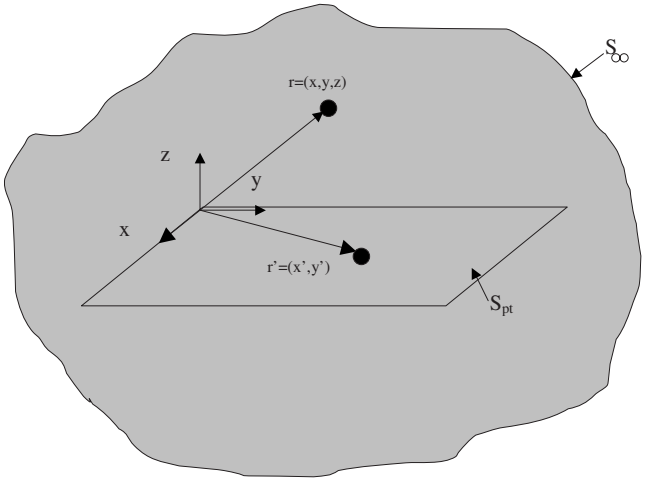


FIG. 2. A schematic of an unbaffled rectangular plate with coordinates (S_∞, infinite surface area of surrounding medium; S_{pt}, total surface area of plate).

Sound power radiated from a simply supported rectangular plate due to the coupled acoustic field is calculated here by using the KH integral formula. Figure 2 shows a simply supported rectangular plate in the infinite medium. The plate dimensions with the corresponding coordinate system are shown in Fig. 1. The sound pressure resulting from the vibration of the plate can be determined from the Helmholtz integral and the Euler equation by applying Green's formula.[21] The Helmholtz equation in rectangular coordinates is given by

$$\nabla^2 p(x,y,z) + k^2 p(x,y,z) = 0. \tag{12}$$

Pressure anywhere in the volume is given by the KH integral[15]

$$P(\vec{r}) = \int_{S_T} \left\{ \frac{\partial G(\vec{r}',\vec{r})}{\partial n_{\vec{r}'}} p(\vec{r}') - \frac{\partial p}{\partial n_{\vec{r}'}} G(\vec{r}',\vec{r}) \right\} dS_T, \tag{13}$$

where vector $\vec{r}$ belongs to volume V, $n_{\vec{r}'}$ is normal to the total surface taken at $\vec{r}'$ comprising S_{pT} and S_∞, pointing outward, $G(\vec{r}',\vec{r})$ is the Green's function, $e^{-jkR}/(4\pi R)$, and

$$R = |\mathbf{r} - \mathbf{r}'| = \sqrt{(x-x')^2 + (y-y')^2 + z^2}. \tag{14}$$

The normal direction depends on which side of the plate is being considered. Accordingly, Eq. (13) may be expanded as

$$P(\vec{r}) = \int_{Sp^+} \left\{ -\frac{\partial G(\vec{r}',\vec{r})}{\partial z_{\vec{r}'}} p^+(\vec{r}') + \frac{\partial p^+}{\partial z_{\vec{r}'}} G(\vec{r}',\vec{r}) \right\} dS_{p^+} + \int_{Sp^-} \left\{ \frac{\partial G(\vec{r}',\vec{r})}{\partial z_{\vec{r}'}} p^-(\vec{r}') - \frac{\partial p^-}{\partial z_{\vec{r}'}} G(\vec{r}',\vec{r}) \right\} dS_{p^-}. \tag{15}$$

Here, $\vec{r} \in V^+$ is half-space located "above" the plate, and $\vec{r}'$ is defined on the surface of the plate, as shown in Fig. 2. At the plate midsurface, S_p, Green's function and acoustic pressure are continuous functions. The simple layer contribution vanishes when S_{p^+} and S_{p^-} tend to S_p. This is reasonably valid for thin plates. So,

$$\int_{Sp} \left\{ \frac{\partial p^+}{\partial z_{\vec{r}'}} - \frac{\partial p^-}{\partial z_{\vec{r}'}} \right\} G(\vec{r}',\vec{r})\} dS_p = 0, \quad \vec{r}' \in S_p. \quad (16)$$

Substituting Eq. (16) into Eq. (15) gives the double layer integral of the pressure

$$P(\vec{r}) = \int_{Sp} \{p^-(\vec{r}') - p^+(\vec{r}')\} \frac{\partial G(\vec{r}',\vec{r})}{\partial z_{\vec{r}'}} dS_p \vec{r} \in V^+,$$

$$\vec{r}' \in S_p. \quad (17)$$

Substituting the pressure jump $\overline{\mathbf{P}}(\vec{r}') = p^-(\vec{r}') - p^+(\vec{r}')$ into Euler's equation yields

$$w(\vec{r}) = \frac{j}{\rho_0 \omega} \int_{Sp} \overline{\mathbf{P}}(\vec{r}') \frac{\partial^2 G(\vec{r}',\vec{r})}{\partial z_{\vec{r}'} \partial z_{\vec{r}}} dS_p. \quad (18)$$

This equation is solved for pressure jump by means of the variational approach. The surface acoustic pressure jump is expressed in terms of the acoustic admissible functions (φ_{m_1,m_2}) as

$$\overline{\mathbf{P}}(x,y) = \sum_{m_1 m_2} \overline{\mathbf{P}}_{m_1 m_2} \varphi_{m_1 m_2}(x,y). \quad (19)$$

The stationary condition for the fluid functional[15] gives the following relation between plate deflection and pressure jump in terms of admittance:

$$\frac{1}{\omega^2} \sum_{m_1' m_2'} A^p_{m_1 m_2 m_1' m_2'} \bar{P}_{m_1' m_2'} + \sum_{ij} S_{m_1 m_2 ij} w_{ij} = 0, \quad (20)$$

where

$$\mathbf{A^p} = -\omega^2 \mathbf{M^A} + \mathbf{K^A} \quad (21)$$

with

$$M^A_{m_1 m_2 m_1' m_2'} = \frac{1}{\rho_0 c^2} \int \int_{S_p \times S_p} \int \int \varphi_{m_1 m_2}(\vec{r}) \times G(\vec{r},\vec{r}') \varphi_{m_1' m_2'}(\vec{r}) dS dS' \quad (22)$$

and

$$K^A_{m_1 m_2 m_1' m_2'} = \frac{1}{\rho_0} \int \int_{S_p \times S_p} \int \int \left\{ \frac{\partial \varphi_{m_1 m_2}(\vec{r})}{\partial x} \frac{\partial \varphi_{m_1' m_2'}(\vec{r})}{\partial x'} + \frac{\partial \varphi_{m_1 m_2}(\vec{r})}{\partial y} \frac{\partial \varphi_{m_1' m_2'}(\vec{r})}{\partial y'} \right\} G(\vec{r},\vec{r}') dS dS'. \quad (23)$$

We define $\mathbf{S}$ as a transform from a plate vibration mode shape to an acoustic pressure jump admissible function as

$$S_{m_1 m_2 m_1' m_2'} = \int \int_{S_p} \varphi_{m_1 m_2}(\vec{r}') \varphi_{m_1' m_2'}(\vec{r}') dS. \quad (24)$$

Equation (20) yields the acoustic pressure jump value for the UBP as follows:

$$\overline{\mathbf{P}}(x,y) = -\omega^2 \sum_{ij} \left(\sum_{kl} \sum_{m_1' m_2'} A^{-1}_{ijkl} S_{klpq} w_{m_1' m_2'} \right) \varphi_{ij}(x,y). \quad (25)$$

The above equation is expressed in another form similar to that for the BP radiation impedance[15] as follows:

$$\overline{\mathbf{P}}(\mathbf{x},\mathbf{y}) = j\omega \sum_{m_1' m_2'} (Z^{\mathrm{UBP}}_{m_1 m_2 m_1' m_2'} w_{m_1' m_2'}), \quad (26)$$

where

$$Z^{\mathrm{UBP}}_{m_1 m_2 m_1' m_2'} = j\omega \sum_{ij} \sum_{kl} S^t_{m_1 m_2 ij} A^{-1}_{ijkl} S_{klm_1' m_2'}$$

or in the matrix form

$$\mathbf{Z}^{\mathrm{UBP}} = j\omega \mathbf{S^t A^{-1} S}, \quad (27)$$

where $\mathbf{S^t}$ is transpose of the matrix $\mathbf{S}$.

The radiated sound power can be obtained by integrating the acoustic intensity over the surface of the plate and it gives sound power radiation from both sides of the plate. The sound power radiation in only the outside direction of cavity is assumed to be as half as much. Of course, this is valid for noise radiation from an isolated plate or wall only. Thus,

$$W_{\mathrm{rad}} = \int_0^{L_2} \int_0^{L_1} \frac{1}{2} \mathrm{Re}[w^*(\mathbf{r}) \overline{\mathbf{P}}(\mathbf{r})|_{z=0}] dx dy, \quad (28)$$

where * denotes the complex conjugate of a complex number. Substituting the plate vibration (1b) and the acoustic pressure jump (26) into the acoustic power (28) gives sound power radiation from the plate in terms of the radiation impedance and complex amplitude of the vibration velocity modes as follows:

$$W_{\mathrm{rad}} = \tfrac{1}{2} \mathbf{b}^H \mathbf{Re}[\mathbf{Z}^{\mathrm{UBP}}] \mathbf{b}. \quad (29)$$

The superscript $\mathbf{H}$ denotes the Hermitian of a matrix, and real part of the specific radiation impedance is calculated from the matrix $\mathbf{A^p}$.

IV. CALCULATION OF THE $\mathbf{A^p}$ MATRIX

The coefficients of matrix $\mathbf{A}^{\mathrm{p}}$ are calculated from the coefficients of matrices $M_{m_1 m_2 m_1' m_2'}$ and $K_{m_1 m_2 m_1' m_2'}$. Substituting the mode shape function for a simply supported rectangular plate as defined in Eq. (3) into Eqs. (22) and (23) gives the expressions for $M_{m_1 m_2 m_1' m_2'}$ and $K_{m_1 m_2 m_1' m_2'}$ as follows:

$$M^{\mathbf{A}} = \frac{1}{\pi \rho_0 c^2} \int_0^{L_2} \int_0^{L_1} \int_0^{L_2} \int_0^{L_1} \sin\left(\frac{m_1 \pi x}{L_1}\right) \sin\left(\frac{m_2 \pi y}{L_2}\right) \times \sin\left(\frac{m_1' \pi x'}{L_1}\right) \sin\left(\frac{m_2' \pi y'}{L_2}\right) \frac{e^{-jkR}}{R} dx' dy' dx dy, \quad (30)$$

$$\mathrm{K^A} = \frac{1}{\pi\rho_0}\int_0^{L_2}\int_0^{L_1}\int_0^{L_2}\int_0^{L_1} \times\left\{\cos\left(\frac{m_1\pi x}{L_1}\right)\cos\left(\frac{m_1'\pi x'}{L_1}\right)\sin\left(\frac{m_2\pi y}{L_2}\right) \times\sin\left(\frac{m_2'\pi y'}{L_2}\right)\frac{m_1m_1'\pi^2}{L_1^2} + \sin\left(\frac{m_1\pi x}{L_1}\right)\sin\left(\frac{m_1'\pi x'}{L_1}\right) \times\cos\left(\frac{m_2\pi y}{L_2}\right)\cos\left(\frac{m_2'\pi y'}{L_2}\right)\frac{m_2m_2'\pi^2}{L_2^2}\right\} \times\frac{e^{-jkR}}{R}dx'dy'dxdy. \quad (31)$$

It may be noted that the expression for $\mathrm{Im}[M^A]$ is the same as Eq. (7) in Ref. 14. $\mathbf{K^A}$ may be expanded as follows:

$$\mathbf{K^A} = \mathbf{K_{11}^A} - \mathbf{jK_{12}^A} + \mathbf{K_{21}^A} - \mathbf{jK_{22}^A}, \quad (32)$$

$$\mathbf{K_{12}^A} = \frac{m_1m_1'\pi}{\rho_0L_1^2}\int_0^{L_2}\int_0^{L_1}\int_0^{L_2}\int_0^{L_1} \times\left\{\cos\left(\frac{m_1\pi x}{L_1}\right)\cos\left(\frac{m_1'\pi x'}{L_1}\right)\sin\left(\frac{m_2\pi y}{L_2}\right) \times\sin\left(\frac{m_2'\pi y'}{L_2}\right)\frac{\sin(kR)}{R}dx'dy'\right\}dxdy, \quad (33a)$$

$$\mathbf{K_{22}^A} = \frac{m_2m_2'\pi}{\rho_0L_2^2}\int_0^{L_2}\int_0^{L_1}\int_0^{L_2}\int_0^{L_1} \times\left\{\sin\left(\frac{m_1\pi x}{L_1}\right)\sin\left(\frac{m_1'\pi x'}{L_1}\right)\cos\left(\frac{m_2\pi y}{L_2}\right) \times\cos\left(\frac{m_2'\pi y'}{L_2}\right)\frac{\sin(kR)}{R}\right\}dx'dy'. \quad (33b)$$

$\mathbf{K_{11}^A}$ and $\mathbf{K_{21}^A}$ may be expressed similar to $\mathbf{K_{12}^A}$ and $\mathbf{K_{22}^A}$ by the replacing $\sin(kR)$ term with $\cos(kR)$, respectively. Quadruple integrals in Eqs. (30) and (31) are evaluated by means of the established method of Refs. 14 and 22. This method considers the self-radiation terms (individual mode contribution in the sound radiation) and the mutual radiation terms (mode coupling contribution in sound radiation) without any major assumptions. The first step in this method is to convert the quadruple integral into double integrals by a coordinate transformation, and the next step is to convert these double integrals into single integrals by using polar coordinates. The self- and mutual radiation terms can be expressed as follows.

Self radiation terms: For $m_1=m_1'$ and $m_2=m_2'$,

$$\mathrm{Im}(M_{m_1m_2,m_1m_2}) = \frac{1}{\pi\rho_0c^2}\left\{\frac{1}{\alpha_{m_1}\beta_{m_2}}J_1^{m_1m_2} + J_2^{m_1m_2} + \frac{1}{\alpha_{m_1}}J_3^{m_1m_2} + \frac{1}{\beta_{m_2}}J_4^{m_1m_2}\right\},$$

$$K_{m_1m_2,m_1m_2}^{12} = \frac{m_1^2\pi}{\rho_0L_1^2}\left\{-\frac{1}{\alpha_{m_1}\beta_{m_2}}J_1^{m_1m_2} + J_2^{m_1m_2} - \frac{1}{\alpha_{m_1}}J_3^{m_1m_2} + \frac{1}{\beta_{m_2}}J_4^{m_1m_2}\right\},$$

$$K_{m_1m_2,m_1m_2}^{22} = \frac{m_2^2\pi}{\rho_0L_2^2}\left\{-\frac{1}{\alpha_{m_1}\beta_{m_2}}J_1^{m_1m_2} + J_2^{m_1m_2} + \frac{1}{\alpha_{m_1}}J_3^{m_1m_2} - \frac{1}{\beta_{m_2}}J_4^{m_1m_2}\right\}. \quad (34)$$

Mutual radiation terms: For $m_1\neq m_1'$ and $m_2\neq m_2'$,

$$\mathrm{Im}(M_{m_1m_2,m_1'm_2'}) = \frac{\varepsilon(m_1'-m_1)\varepsilon(m_2'-m_2)}{(\alpha_{m_1}^2-\alpha_{m_1'}^2)(\beta_{m_2}^2-\beta_{m_2'}^2)}\{\alpha_{m_1}\beta_{m_2}J_1^{m_1'm_2'} - \alpha_{m_1}\beta_{m_2'}J_1^{m_1'm_2} - \alpha_{m_1'}\beta_{m_2}J_1^{m_1m_2'} + \alpha_{m_1'}\beta_{m_2'}J_1^{m_1m_2}\},$$

$$K_{m_1m_2,m_1'm_2'}^{12} = \frac{m_1m'\pi}{\rho_0L_1^2}\frac{\varepsilon(m_1'-m_1)\varepsilon(m_2'-m_2)}{(\alpha_{m_1}^2-\alpha_{m_1'}^2)(\beta_{m_2}^2-\beta_{m_2'}^2)} \times\{\alpha_{m_1'}\beta_{m_2}J_1^{m_1'm_2'} - \alpha_{m_1'}\beta_{m_2'}J_1^{m_1'm_2} - \alpha_{m_1}\beta_{m_2}J_1^{m_1m_2'} + \alpha_{m_1}\beta_{m_2'}J_1^{m_1m_2}\},$$

$$K_{m_1m_2,m_1'm_1'}^{22} = \frac{m_2m_2'\pi}{\rho_0L_2^2}\frac{\varepsilon(m_1'-m_1)\varepsilon(m_2'-m_2)}{(\alpha_{m_1}^2-\alpha_{m_1'}^2)(\beta_{m_2}^2-\beta_{m_2'}^2)} \times\{\alpha_{m_1}\beta_{m_2'}J_1^{m_1'm_2'} - \alpha_{m_1}\beta_{m_2}J_1^{m_1'm_2} - \alpha_{m_1'}\beta_{m_2'}J_1^{m_1m_2'} + \alpha_{m_1'}\beta_{m_2}J_1^{m_1m_2}\}. \quad (35)$$

For $m_1\neq m_1'$ and $m_2=m_2'$,

$$\mathrm{Im}(M_{m_1m_2,m_1'm_2}) = \frac{\varepsilon(m_1'-m_1)}{(\alpha_{m_1}^2-\alpha_{m_1'}^2)}\left\{\alpha_{m_1}J_3^{m_1'm_2} - \alpha_{m_1'}J_3^{m_1m_2} + \frac{\alpha_{m_1}}{\beta_{m_2}}J_1^{m_1'm_2} - \frac{\alpha_{m_1'}}{\beta_{m_2}}J_1^{m_1m_2}\right\},$$

$$K_{m_1m_2,m_1'm_2}^{12} = \frac{m_1m_1'\pi}{\rho_0L_1^2}\frac{\varepsilon(m_1'-m_1)}{(\alpha_{m_1}^2-\alpha_{m_1'}^2)}\left\{\alpha_{m_1'}J_3^{m_1'm_2} - \alpha_{m_1}J_{3i}^{m_1m_2} + \frac{\alpha_{m_1'}}{\beta_{m_2}}J_{1i}^{m_1'm_2} - \frac{\alpha_{m_1}}{\beta_{m_2}}J_{1i}^{m_1m_2}\right\},$$

$$K^{22}_{m_1m_2,m_1'm_2} = \frac{m_2m_2'\pi}{\rho_0 L_2^2}\frac{\varepsilon(m_1'-m_1)}{(\alpha_{m_1}^2-\alpha_{m_1'}^2)}\left\{\alpha_{m_1}J_3^{m_1'm_2}-\alpha_{m_1'}J_3^{m_1m_2}+\frac{\alpha_{m_1'}}{\beta_{m_2}}J_1^{m_1m_2}-\frac{\alpha_{m_1}}{\beta_{m_2}}J_1^{m_1'm_2}\right\}. \quad (36)$$

For $m_1=m_1'$ and $m_2\neq m_2'$,

$$\mathrm{Im}(M_{m_1m_2,m_1m_2'}) = \frac{\varepsilon(m_2'-m_2)}{(\beta_{m_2}^2-\beta_{m_2'}^2)}\left\{\beta_{m_2}J_4^{m_1m_2'}-\beta_{m_2'}J_4^{m_1m_2}+\frac{\beta_{m_2}}{\alpha_{m_1}}J_1^{m_1m_2'}-\frac{\beta_{m_2'}}{\alpha_{m_1}}J_1^{m_1m_2}\right\},$$

$$K^{12}_{m_1m_2,m_1m_2'} = \frac{m_1m_1'\pi}{\rho_0 L_1^2}\frac{\varepsilon(m_2'-m_2)}{(\beta_{m_2}^2-\beta_{m_2'}^2)}\left\{\beta_{m_2}J_4^{m_1m_2'}-\beta_{m_2'}J_4^{m_1m_2}-\frac{\beta_{m_2}}{\alpha_{m_1}}J_1^{m_1m_2'}+\frac{\beta_{m_2'}}{\alpha_{m_1}}J_1^{m_1m_2}\right\},$$

$$K^{22}_{m_1m_2,m_1m_2'} = \frac{m_2m_2'\pi}{\rho_0 L_2^2}\frac{\varepsilon(m_2'-m_2)}{(\beta_{m_2}^2-\beta_{m_2'}^2)}\left\{\beta_{m_2'}J_4^{m_1m_2'}-\beta_{m_2}J_4^{m_1m_2}+\frac{\beta_{m_2'}}{\alpha_{m_1}}J_1^{m_1m_2'}-\frac{\beta_{m_2}}{\alpha_{m_1}}J_1^{m_1m_2}\right\}, \quad (37)$$

where

$$\begin{Bmatrix} J_{1i}^{m_1m_2} \\ J_{2i}^{m_1m_2} \\ J_{3i}^{m_1m_2} \\ J_{4i}^{m_1m_2} \end{Bmatrix} = \int_0^{L_2}\int_0^{L_1}\begin{Bmatrix} 1 \\ (L_1-x)(L_2-y) \\ (L_2-y) \\ (L_1-x) \end{Bmatrix} \times \begin{Bmatrix} \sin\alpha_{m_1}x\sin\beta_{m_2}y \\ \cos\alpha_{m_1}x\cos\beta_{m_2}y \\ \sin\alpha_{m_1}x\cos\beta_{m_2}y \\ \cos\alpha_{m_1}x\sin\beta_{m_2}y \end{Bmatrix} \times \frac{\cos(k\sqrt{x^2+y^2})}{\sqrt{x^2+y^2}}dxdy, \quad (38)$$

$$\alpha_{m_1} = \frac{m_1\pi}{L_1}, \quad \beta_{m_2} = \frac{m_2\pi}{L_2}, \quad \varepsilon(m_1'-m_1) = \begin{cases} 1 & m_1' = m_1 \\ 0 & \text{for } m_1'-m_1 = \pm 1, \pm 3, \pm 5, \ldots \\ 2 & m_1'-m_1 = \pm 2, \pm 4, \pm 6, \ldots . \end{cases} \quad (39)$$

$\mathrm{Re}(\mathbf{M^A})$, $K^{11}_{m_1m_2,m_1m_2'}$, and $K^{21}_{m_1m_2,m_1m_2'}$ terms are obtained by replacing J_1–J_4 in the $\mathrm{Im}(\mathbf{M^A})$, $K^{12}_{m_1m_2,m_1m_2'}$, and $K^{22}_{m_1m_2,m_1m_2'}$ with J_{1i}–J_{4i}, respectively. $J_1^{m_1m_2}, J_2^{m_1m_2}, J_3^{m_1m_2}$ and $J_4^{m_1m_2}$ may be expressed as in Eq. (38) above by replacing $\cos(k\sqrt{x^2+y^2})$ with $\sin(k\sqrt{x^2+y^2})$. The double integrals involved in Eq. (38) are simplified in terms of single integrals by using polar coordinate transformation as follows.[23]

$$\begin{Bmatrix} x \\ y \end{Bmatrix} = L_1\rho\begin{Bmatrix} \cos(\theta) \\ \sin(\theta) \end{Bmatrix}. \quad (40)$$

The double integrals involved in Eq. (38) can be written as

$$\begin{aligned} \mathfrak{I}_{1i}^{m_1m_2} &= \int_0^{L_2}\int_0^{L_1}\sin\alpha_{m_1}x\sin\beta_{m_2}y\frac{\cos k\sqrt{x^2+y^2}}{\sqrt{x^2+y^2}}dxdy \\ &= \mathfrak{I}_{1i}^{(1)}(\alpha_{m_1},\beta_{m_2},L_1,L_2)+\mathfrak{I}_{1i}^{(2)}(\alpha_{m_1},\beta_{m_2},L_1,L_2), \\ \mathfrak{I}_{2i}^{m_1m_2} &= \int_0^{L_2}\int_0^{L_1}\cos\alpha_{m_1}x\cos\beta_{m_2}y\frac{\cos k\sqrt{x^2+y^2}}{\sqrt{x^2+y^2}}dxdy \\ &= \mathfrak{I}_{2i}^{(1)}(\alpha_{m_1},\beta_{m_2},L_1,L_2)+\mathfrak{I}_{2i}^{(2)}(\alpha_{m_1},\beta_{m_2},L_1,L_2), \\ \mathfrak{I}_{3i}^{m_1m_2} &= \int_0^{L_2}\int_0^{L_1}\sin\alpha_{m_1}x\cos\beta_{m_2}y\frac{\cos k\sqrt{x^2+y^2}}{\sqrt{x^2+y^2}}dxdy \\ &= \mathfrak{I}_{3i}^{(1)}(\alpha_{m_1},\beta_{m_2},L_1,L_2)+\mathfrak{I}_{3i}^{(2)}(\alpha_{m_1},\beta_{m_2},L_1,L_2), \\ \mathfrak{I}_{4i}^{m_1m_2} &= \int_0^{L_2}\int_0^{L_1}\cos\alpha_{m_1}x\sin\beta_{m_2}y\frac{\cos k\sqrt{x^2+y^2}}{\sqrt{x^2+y^2}}dxdy \\ &= \mathfrak{I}_{4i}^{(1)}(\alpha_{m_1},\beta_{m_2},L_1,L_2)+\mathfrak{I}_{4i}^{(2)}(\alpha_{m_1},\beta_{m_2},L_1,L_2), \end{aligned} \quad (41)$$

where

$$\mathfrak{I}_{1i}^{(1)} = \int_0^{\tan^{-1}(\mu)}\int_0^{\sec(\theta)}\sin(L_1\alpha_{m_1}\rho\cos\theta)\times\sin(L_1\beta_{m_2}\rho\sin\theta)\cos(kL_1\rho)L_1d\rho d\theta, \quad (42a)$$

$$\mathfrak{I}_{1i}^{(2)} = \int_{\tan^{-1}(\mu)}^{\pi/2}\int_0^{\mu\csc(\theta)}\sin(L_1\alpha_{m_1}\rho\cos\theta)\times\sin(L_1\beta_{m_2}\rho\sin\theta)\cos(kL_1\rho)L_1d\rho d\theta, \quad (42b)$$

with $\mu=L_2/L_1$. $\mathfrak{I}_{2i}^{(1)}$, $\mathfrak{I}_{2i}^{(2)}$, $\mathfrak{I}_{3i}^{(1)}$, $\mathfrak{I}_{3i}^{(2)}$, $\mathfrak{I}_{4i}^{(1)}$, and $\mathfrak{I}_{4i}^{(2)}$ expressions are similar to Eqs. (42a) and (42b) with the corresponding trigonometric functions mentioned in Eq. (41). The inner integral of Eqs. (42a) and (42b) can be evaluated analytically. The resulting expressions are given by

$$\mathfrak{I}_{1i}^{(1)} = \frac{1}{4}\int_0^{\tan^{-1}(\mu)}\{R(\alpha_{m_1},-\beta_{m_2},-k,\theta)+R(\alpha_{m_1},-\beta_{m_2},k,\theta)-R(\alpha_{m_1},\beta_{m_2},-k,\theta)-R(\alpha_{m_1},\beta_{m_2},k,\theta)\}d\theta,$$

$$\mathfrak{I}_{2i}^{(1)} = \frac{1}{4}\int_0^{\tan^{-1}(\mu)}\{R(\alpha_{m_1},-\beta_{m_2},-k,\theta)+R(\alpha_{m_1},-\beta_{m_2},k,\theta)+R(\alpha_{m_1},\beta_{m_2},-k,\theta)+R(\alpha_{m_1},\beta_{m_2},k,\theta)\}d\theta,$$

$$\mathfrak{I}_{3i}^{(1)} = \frac{1}{4}\int_0^{\tan^{-1}(\mu)} \{-S(\alpha_{m_1},-\beta_{m_2},-k,\theta) - S(\alpha_{m_1},-\beta_{m_2},k,\theta) - S(\alpha_{m_1},\beta_{m_2},-k,\theta) - S(\alpha_{m_1},\beta_{m_2},k,\theta)\}d\theta,$$

$$\mathfrak{I}_{4i}^{(1)} = \frac{1}{4}\int_0^{\tan^{-1}(\mu)} \{S(\alpha_{m_1},-\beta_{m_2},-k,\theta) + S(\alpha_{m_1},-\beta_{m_2},k,\theta) - S(\alpha_{m_1},\beta_{m_2},-k,\theta) - S(\alpha_{m_1},\beta_{m_2},k,\theta)\}d\theta, \quad (43)$$

where

$$S(\alpha_{m_1},\beta_{m_2},k,\theta) = \frac{\cos(L_1(\alpha_m \cos\theta + \beta \sin\theta + k)\sec\theta)}{(\alpha_m \cos\theta + \beta \sin\theta + k)} - \frac{1}{(\alpha_m \cos\theta + \beta \sin\theta + k)},$$

$$R(\alpha_{m_1},\beta_{m_2},k,\theta) = \frac{\sin(L_1(\alpha_m \cos\theta + \beta \sin\theta + k)\sec\theta)}{(\alpha_m \cos\theta + \beta \sin\theta + k)}. \quad (44)$$

The second part of the integrals in Eq. (41) can be directly obtained from Eq. (43) as follows:

$$\mathfrak{I}_{ji}^{(2)}(\alpha_{m_1},\beta_{m_2},L_1,L_2) = \mathfrak{I}_{ji}^{(1)}(\beta_{m_2},\alpha_{m_1},L_2,L_1) \quad (j=1,2), \quad (45)$$

$$\mathfrak{I}_{3i}^{(2)}(\alpha_{m_1},\beta_{m_2},L_1,L_2) = \mathfrak{I}_{4i}^{(1)}(\beta_{m_2},\alpha_{m_1},L_2,L_1), \quad (46)$$

$$\mathfrak{I}_{4i}^{(2)}(\alpha_{m_1},\beta_{m_2},L_1,L_2) = \mathfrak{I}_{3i}^{(1)}(\beta_{m_2},\alpha_{m_1},L_2,L_1). \quad (47)$$

The integral expressions in Eq. (38) are expressed in terms of double integral values evaluated in Eq. (41) as follows:

$$J_{1i}^{m_1m_2} = \mathfrak{I}_{1i}^{mn}, \quad J_{2i}^{m_1m_2} = L_1L_2\mathfrak{I}_{2i}^{mn} - L_1\frac{\partial}{\partial\beta_{m_2}}\mathfrak{I}_{4i}^{mn} - L_2\frac{\partial}{\partial\beta_{m_2}}\mathfrak{I}_{3i}^{mn} + \frac{\partial^2}{\partial\beta_{m_2}\partial\alpha_{m_1}}\mathfrak{I}_{1i}^{mn},$$

$$J_{3i}^{m_1m_2} = L_2\mathfrak{I}_{3i}^{mn} - \frac{\partial}{\partial\beta_{m_2}}\mathfrak{I}_{1i}^{mn}, \quad J_{4i}^{m_1m_2} = L_1\mathfrak{I}_{4i}^{mn} - \frac{\partial}{\partial\alpha_{m_2}}\mathfrak{I}_{1i}^{mn}. \quad (48)$$

These single integrals are evaluated numerically.

Table I shows a comparison of the computational time to calculate the matrices $\mathbf{M^A}$ and $\mathbf{K^A}$ by using the double integration method of Refs. 15 and 16 and the method proposed here for the single integration. Consider a steel plate of dimensions 1×1 m^2 and thickness 0.9 mm with three arbitrary modes (1, 1), (3, 1), and (3, 3) in the calculation. The matrices have been computed at three different frequencies: 50, 200, and 1000 Hz. The single integral method reduces computational time down to 3%–10%, depending on the frequencies and the number of modes in the calculation. The single integration computational time is independent of the analysis frequency. It has the potential of enormous saving in computational time in practical thin cavity problems where the number of modes would be very large.

TABLE I. Comparison of the computational time to calculate the radiation impedance matrices for a thin steel plate of dimensions (1×1 m^2) with 0.9 mm thickness by double integration and single integration.

		Computational time (s)	
	Frequency (Hz)	Double integral approach	Single integral approach
$\mathbf{M^A}$	50	3.563	0.563
	200	3.891	0.438
	1000	6.75	0.453
$\mathbf{K^A}$	50	7.485	0.454
	200	8.187	0.485
	1000	14.5	0.469

V. CALCULATION OF THE TRANSVERSE TRANSMISSION LOSS

Transverse transmission loss is defined as $\mathrm{TL_{tp}} = 10\log(W_{\mathrm{int}}/W_{\mathrm{rad}})$. The radiated power is calculated by using Eq. (29) in Sec. III. The wave equation in the axial direction (y-direction) is used to calculate the incident power.[24]

$$p(y) = Ae^{-jky} + Be^{jky}; \quad W_{\mathrm{int}} = \frac{|A|^2}{2Y}$$

where

$$A = \frac{(p(0) + \rho_0 c_0 v(0))}{2}, \quad (49)$$

where $Y=\rho_0 c_0/S$ is the characteristic impedance of the cavity for plane waves,[24] $v(0)=s(x,\omega)^*V/S$, $s(x,\omega)$ denotes the acoustic source strength density function (volume velocity for unit volume) in the cavity volume V, and S is the cross-sectional area of cavity.

VI. NUMERICAL MODEL

The finite element/boundary element (FEM-BEM) based numerical model has been used here to validate the analytical method discussed above in order to calculate the transverse transmission loss in the cavity. The FEM-BEM analysis has been performed in the commercial package named SYSNOISE.[25] The procedure used in the analysis of the coupled acoustic structural analysis is linking the "FEM structure" and "FEM fluid" models to simultaneously predict the inside pressure and plate vibration displacement. The vibration displacement is then used as input for "indirect BEM" to predict the radiated power from the flexible plate. Sound source input to the cavity (speaker) is modeled by the equivalent constant velocity source in the SYSNOISE model.

VII. RESULTS AND DISCUSSION

Consider a rectangular cavity of dimensions of $L_1=0.3$ m, $L_2=1.5$ m, and $L_3=0.4$ m, and let the thickness of the flexible aluminum plate be 5 mm. A speaker excites one

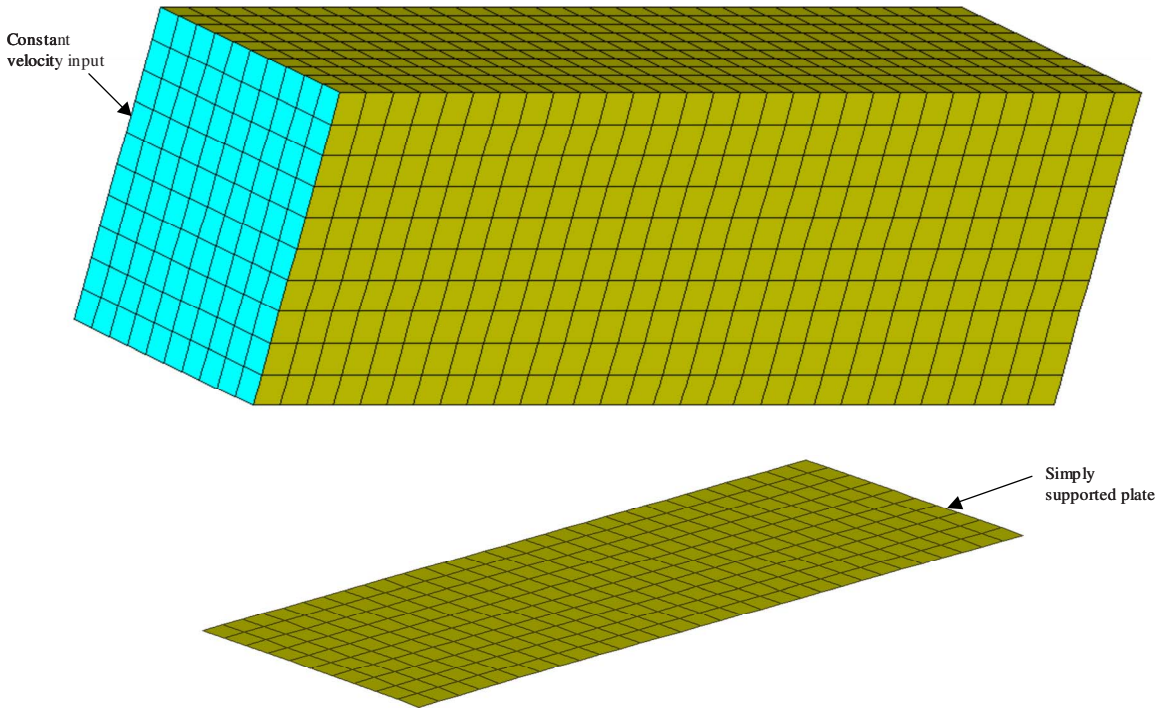


FIG. 3. (Color online) Numerical model of the cavity with constant input velocity source and the unbaffled flexible plate.

end of the cavity acoustically. The material properties used in the analytical calculation are density=2770 kg/m^3, Young's modulus=71×10^9 N/m^2, Poisson's ratio ν=0.33, and the acoustic and structural damping ratio or loss factor=0.01. The air properties are sound speed=340 m/s and density =1.2 kg/m^3. The coupling phenomena and breakout noise are dominant at low frequencies. So, the results are analyzed in the frequency range 10–400 Hz by considering the modes inside the cavity, and the flexible plate up to 800 Hz; that is, double the highest frequency of interest. The cavity dimensions considered in this study are the same as in Ref. 13 and these are called the "cavity 1 dimensions" for further reference in this paper. We have also considered another cavity with dimensions L_1=1.17 m, L_2=0.206 m, and L_3 =0.258 m and wall thickness of 1.295 mm with a zinc coated steel material. These are referred to as the "cavity 2 dimensions."

An FEM fluid model was created in SYSNOISE with 3000 elements of SOLID73 type for a cavity, and a constant velocity boundary condition was applied at one end by calculating the equivalent velocity source strength of the speaker. This model is linked with the "FEM structure" model meshed with 300 elements of the SHELL63 type, and simple boundary conditions are applied. The link option created the coupling between the FEM fluid and the FEM structure and solved for the inside pressure cavity and wall vibration displacement. Figure 3 shows the computational model of the cavity and the panel system.

Figure 4 shows the inside cavity pressure measured at the location $(0.5L_1, 0.4L_2, 0.5L_3)$ in cavity 1 by means of the IMCM model and the SYSNOISE model. Clearly, there is a close agreement between the results of the two methods. It also validates the numerical model dimensions, mesh size, and boundary conditions, and this model is used for further analysis to predict the transverse transmission loss. Table II shows the uncoupled natural frequencies of cavity 1, the panel (plate), and the coupled frequencies of the cavity-panel system. Observation of the calculated results shows that a strong coupling exists between the cavity and the plate in the vicinity of any uncoupled acoustic resonance.

Figure 5 shows the effect of the cavity on calculation of the radiated power from the compliant wall for a given velocity boundary condition. In the low frequency range (below the fundamental mode of the cavity and panel), the radiated power is constant and independent of frequency. Cavity changes the stiffness of the plate, and this affects

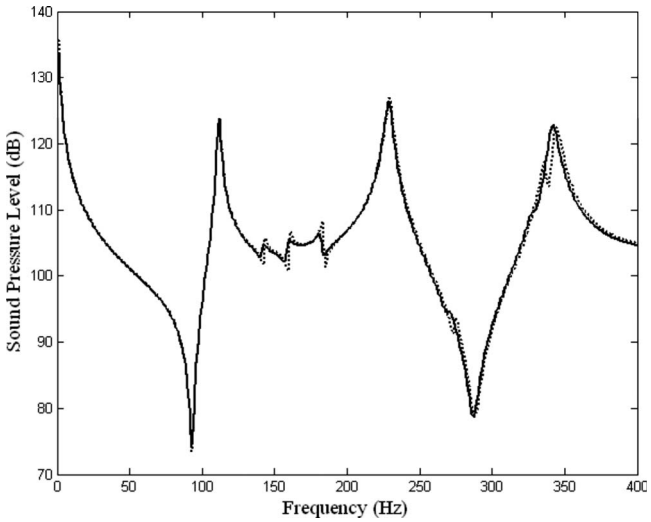


FIG. 4. Comparison of the sound pressure level prediction methodology inside the rectangular cavity at the location of $(0.5L_1, 0.4L_2, 0.5L_3)$ due to acoustic excitation of the structural-acoustic system. (—) Compact matrix method and (- - -) SYSNOISE.

TABLE II. Coupled and uncoupled natural frequencies of the cavity-panel system (cavity dimensions: 0.3 $\times 0.4 \times 1.5$ m^3 with a 5 mm aluminum flexible cavity wall).

Coupled modes		Acoustic modes		Structural modes	
Mode No.	Frequency (Hz)	Mode No.	Frequency (Hz)	Mode No.	Frequency (Hz)
1	111.6	1	113.4	1	142.1
2	142.9	2	227.1	2	158.2
3	160.0	3	341.4	3	185.5
4	183.5	4	426.8	4	224.1
5	223.4	5	441.6	5	275.0
6	229.1	6	456.6	6	338.8
7	274.6	7	483.4	7	416.7
8	336.4				
9	344.1				
10	415.8				
11	428.5				
12	442.8				
13	457.5				
14	484.7				

noise radiation in the stiffness-controlled region. This is not captured by the analytical model presented here. The plane wave is there inside the cavity at all frequencies and this is used for the structure to resonate at its (structural) resonances and radiate noise outside. The net result is as follows.

(i) Structural resonances dominate the radiated power.
(ii) If an internal (acoustic) natural frequency nearly coincides with any structural natural frequency, then the radiated power peak will be particularly sharp.
(iii) Structural damping is therefore crucial in reducing breakout noise.

Analytical model of the unbaffled compliant wall when compared with the SYSNOISE model of the cavity-backed compliant wall shows that a major discrepancy in sound power radiation is in the very low frequency region (below the fundamental mode of the cavity and the panel). This is because the stiffness effect of the air enclosed in a cavity (or box) is not captured by the analytical model presented here. Incidentally, the BP model is no better than the UBP model, as shown in Fig. 5. This is indeed a weakness of the analytical model presented here. Fortunately, however, at such low frequencies (below 100 Hz) the radiated sound power level is 20–40 dB lower than that in the midfrequency range (100–400 Hz), and therefore is of not much practical significance.

Figure 6 shows the transverse transmission loss (TL_{tp}) calculation for the isolated cavity walls of the cavity 1 dimensions. The predicted values of TL_{tp} may be seen to tally closely with those of the FEM-BEM of SYSNOISE.

Figures 7 and 8 show the effect of coupling on the inside cavity pressure for dimensions of cavity 1 and cavity 2, respectively. Equations (6a) and (6b) are modified for weak coupling by approximating $(\mathbf{I}+\mathbf{Z_a C Y_s C^T}) \cong \mathbf{I}$ and $(\mathbf{I}+\mathbf{Y_s C^T Z_a C}) \cong \mathbf{I}$. The coupling has a minimal effect in cavity 1 as compared to cavity 2, due to larger thickness of the flexible wall in the case of cavity 1. Generally, coupling is stronger for thin cavity walls compared to thick cavity walls.

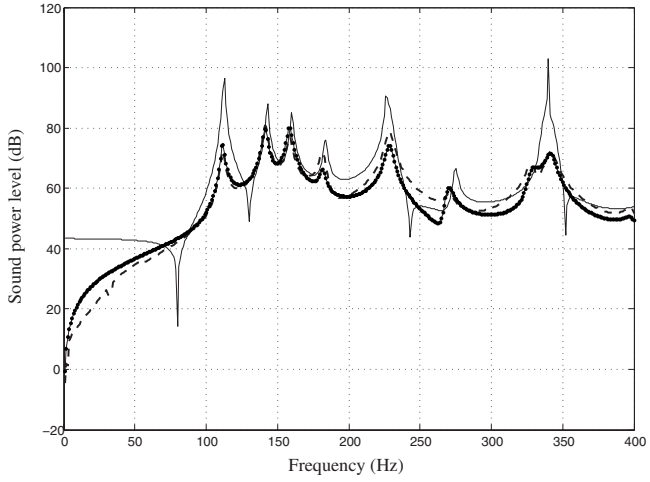


FIG. 5. Effect of cavity on the sound power radiation from the flexible cavity wall (cavity dimensions of $0.3 \times 0.4 \times 1.5$ m^3, and 5 mm flexible cavity wall thickness of 5 mm), due to acoustic excitation of the structural-acoustic system. (—) SYSNOISE model of the cavity-backed plate, (- - -) analytical model of the UBP, and (-●-) analytical model of the BP.

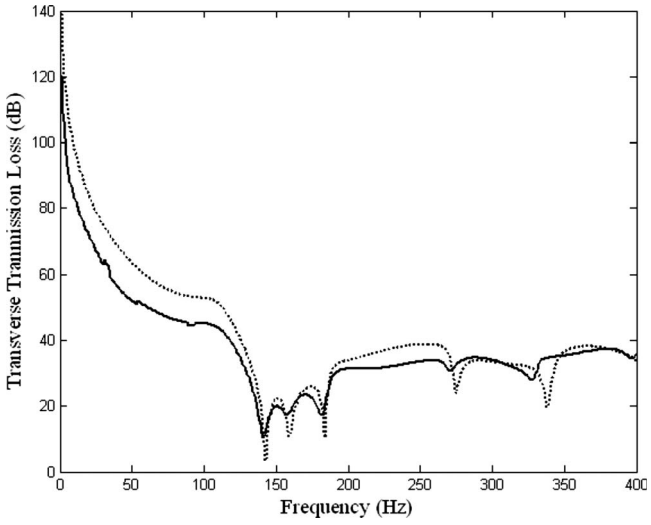


FIG. 6. Prediction of the transverse transmission loss from the flexible simply supported wall of the rectangular cavity 1 dimensions. (—) analytical and (- - -) SYSNOISE.

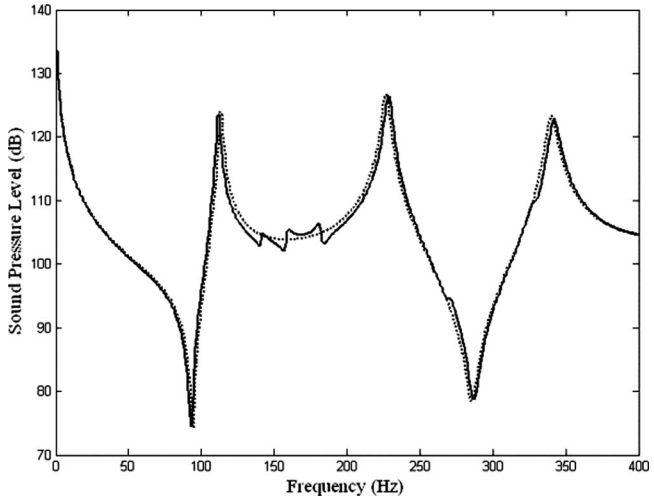


FIG. 7. Effect of coupling on the sound pressure level prediction inside a rectangular cavity of dimensions ($0.3\times0.4\times1.5$ m^3) with a 5 mm aluminum flexible cavity wall, due to acoustic excitation of the structural-acoustic system. (- - -) Uncoupled and (—) coupled.

Figure 8 shows the coupling effect in the region of 140–190 Hz due to the well-coupled modes of structure and cavity. The transfer factor is used to identify the well-coupled modes.[26] The transfer factor for thin cavity flexible wall for the nth cavity mode and mth structural mode can be written as

$$T_{n,m}=\left(1+\frac{(\omega_n^2-\omega_m^2)\rho_s h S_f V}{4\rho_0 c_0^2 C_{n,m}^2}\right)^{-1}. \tag{50}$$

If $|T_{n,m}|\cong 1$, then the cavity mode and the structural mode are well coupled and the energy involved in the interaction is comparable with that of the constituent subsystems, However, if $|T_{n,m}|\ll 1$, then the modal interaction is small and the system is said to be weakly coupled. The possibilities of coupling based on the transfer factor of Eq. (50) for cavity 1 dimensions are acoustic modes 2, 3, and 4 with structural modes 2, 4, and 6 at the cavity natural frequencies of 113, 227, and 340 Hz, and with flexible panel natural frequencies of 157, 222, and 330 Hz, respectively.

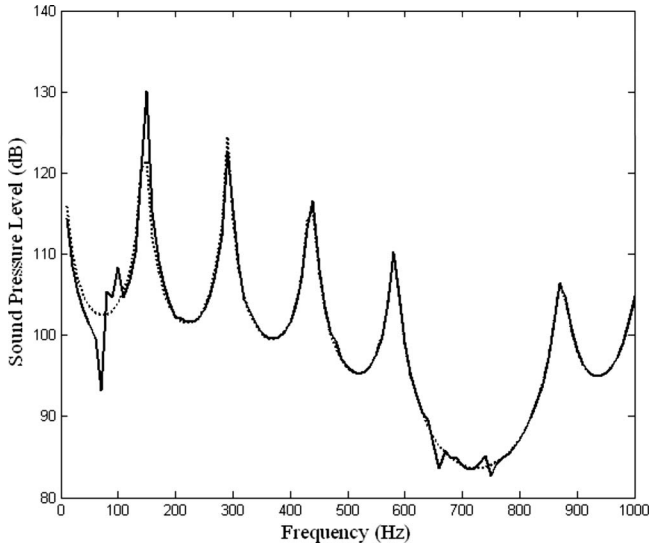


FIG. 8. Effect of coupling on the sound pressure level prediction inside a rectangular cavity of dimensions ($0.203\times0.258\times1.17$ m^3) with a 1.295 mm zinc coated steel flexible cavity wall, due to acoustic excitation of the structural-acoustic system. (- - -) Uncoupled and (—) coupled.

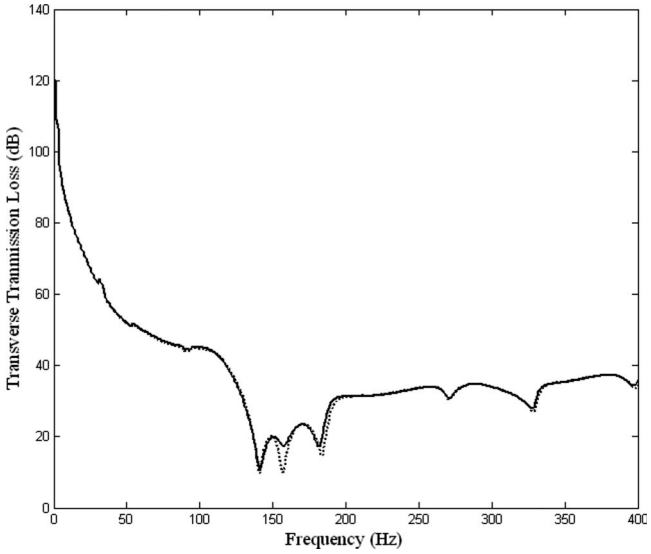


FIG. 9. Effect of coupling on the transverse transmission loss prediction of a rectangular cavity of dimensions ($0.3\times0.4\times1.5$ m^3) with a 5 mm aluminum flexible cavity wall, due to acoustic excitation of the structural-acoustic system. (—) Coupled and (- - -) uncoupled.

Figure 9 shows the effect of coupling on the transverse transmission loss calculation. It may be noted that the effect of coupling is confined to a very narrow frequency range at low frequencies. If a strong coupling occurs between a cavity mode and a panel mode, then damping of the subsystem would control it.

VIII. CONCLUSIONS

An analytical method is proposed in this paper to predict the breakout noise from a rectangular cavity, incorporating three-dimensional effects, by combining the IMCM approach with the KH integral. This method is validated by means of the numerical results obtained from SYSNOISE models as also those available in literature.

A computationally efficient method is suggested to predict the sound power radiated from a thin flexible plate by converting the quadruple integral into single integrals by using coordinate transformation in two steps. It has been shown that this method reduces the computational times drastically. Besides, it has the potentiality of enormous saving in computational time in practical thin cavity problems where the number of modes would be very large.

The assumption of unbaffled compliant wall sound radiation model has a limitation inasmuch as it does not hold in the very low frequency region (below the fundamental mode of cavity and complaint wall). Besides, it has been assumed here that the acoustic power output is divided equally between the outside and inside of the cavity. This may hold for an acoustically lined cavity only.

The next step will be to develop the subsystem level models for different cavity boundary conditions, and then to

extend the current methodology of the breakout noise prediction to the four-side flexible-wall ducts typical of the HVAC systems.

ACKNOWLEDGMENTS

The authors would like to thank GE Global Research, Bangalore and the Department of Science and Technology of the Government of India for providing the required resources for the research work reported here.

[1]E. H. Dowell and H. M. Voss, "The effect of a cavity on panel vibrations," AIAA J. **1**, 476–477 (1963).
[2]A. J. Pretlove, "Free vibrations of a rectangular panel backed by a closed rectangular cavity," J. Sound Vib. **2**, 197–206 (1965).
[3]A. J. Pretlove, "Forced vibrations of a rectangular panel backed by a closed rectangular cavity," J. Sound Vib. **3**, 252–261 (1966).
[4]A. J. Pretlove, "Acoustic-elastic effect in the response of large windows to sonic bangs," J. Sound Vib. **9**, 487–500 (1969).
[5]J. Pan, C. H. Hansen, and D. A. Bies, "Active control of noise transmission through a panel into a cavity: I. Analytical study," J. Acoust. Soc. Am. **87**, 2098–2108 (1990).
[6]J. Pan, "The forced response of an acoustic-structural coupled system," J. Acoust. Soc. Am. **91**, 949–956 (1992).
[7]J. Pan and D. A. Bies, "The effect of fluid-structural coupling on sound waves in an enclosure: Theoretical part," J. Acoust. Soc. Am. **87**, 691–707 (1990).
[8]R. W. Guy, "The response of a cavity backed panel to external airborne excitation: A general analysis," J. Acoust. Soc. Am. **65**, 719–731 (1979).
[9]R. W. Guy and M. C. Bhattacharya, "The transmission of sound through a cavity backed finite plate," J. Sound Vib. **27**, 207–223 (1973).
[10]M. I. Qaisi, "Free vibration of a rectangular plate-cavity system," Appl. Acoust. **24**, 49–61 (1988).
[11]R. H. Lyon, "Noise reduction of rectangular enclosures with one flexible wall," J. Acoust. Soc. Am. **35**, 1791–1797 (1963).
[12]E. H. Dowell, G. F. Gorman and D. A. Smith, "Acoustoelasticity: General theory, acoustic modes and forced response to sinusoidal excitation, including comparisons with experiment," J. Sound Vib. **52**, 519–542 (1977).
[13]S. M. Kim and M. J. Brennan, "A compact matrix formulation using the impedance and mobility approach for the analysis of structural-acoustic systems," J. Sound Vib. **223**, 97–113 (1999).
[14]W. L. Li and H. J. Gilbeling, "Determination of the mutual radiation resistance of a rectangular plate and their impact on the radiated sound power," J. Sound Vib. **229**, 1213–1233 (2000).
[15]H. Nelisse, O. Beslin, and J. Nicolas, "A generalized approach for the acoustic radiation from a baffled or unbaffled plate with arbitrary boundary conditions, immersed in a light or heavy fluid," J. Sound Vib. **211**, 207–225 (1998).
[16]H. Nelisse, O. Beslin, and J. Nicolas, "Fluid-structure coupling for an unbaffled elastic panel immersed in a diffuse field," J. Sound Vib. **198**, 485–506 (1996).
[17]A. J. Bullmore, P. A. Nelson, A. R. D. Curtis, and S. J. Elliott, "The active minimization of harmonic enclosed sound fields, part II: A computer simulation," J. Sound Vib. **117**, 15–33 (1987).
[18]P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).
[19]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[20]S. D. Snyder and N. Tanaka, "On feed forward active control of sound and vibration using vibration error signals," J. Acoust. Soc. Am. **94**, 2181–2193 (1993).
[21]F. Fahy, *Sound and Structural Vibration* (Academic, London, 1989).
[22]W. L. Li and H. J. Gilbeling, "Acoustic radiation from a rectangular plate reinforced by finite springs at arbitrary location," J. Sound Vib. **220**, 117–133 (1999).
[23]E. G. Williams, "A series expansion of the acoustic power radiated from planar sources," J. Sound Vib. **73**, 1520–1524 (1983).
[24]M. L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley, New York, 1987).
[25]SYSNOISE Rev 5.6, Users Manual, LMS International, 2003.
[26]W. H. Louisell, *Coupled Mode and Parametric Electronics* (Wiley, New York, 1960).

Signal denoising and ultrasonic flaw detection via overcomplete and sparse representations

Guang-Ming Zhang and David M. Harvey
General Engineering Research Institute, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, United Kingdom

Derek R. Braden
Delphi Electronics and Safety, Kirkby, Liverpool L33 7XL, United Kingdom

(Received 28 April 2008; revised 19 August 2008; accepted 19 August 2008)

Sparse signal representations from overcomplete dictionaries are the most recent technique in the signal processing community. Applications of this technique extend into many fields. In this paper, this technique is utilized to cope with ultrasonic flaw detection and noise suppression problem. In particular, a noisy ultrasonic signal is decomposed into sparse representations using a sparse Bayesian learning algorithm and an overcomplete dictionary customized from a Gabor dictionary by incorporating some *a priori* information of the transducer used. Nonlinear postprocessing including thresholding and pruning is then applied to the decomposed coefficients to reduce the noise contribution and extract the flaw information. Because of the high compact essence of sparse representations, flaw echoes are packed into a few significant coefficients, and noise energy is likely scattered all over the dictionary atoms, generating insignificant coefficients. This property greatly increases the efficiency of the pruning and thresholding operations and is extremely useful for detecting flaw echoes embedded in background noise. The performance of the proposed approach is verified experimentally and compared with the wavelet transform signal processor. Experimental results to detect ultrasonic flaw echoes contaminated by white Gaussian additive noise or correlated noise are presented in the paper. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2982414]

PACS number(s): 43.60.Hj, 43.60.Wy, 43.60.Mn [EJS] Pages: 2963–2972

I. INTRODUCTION

Ultrasonic inspection is one of the widely used techniques for nondestructive evaluation (NDE) of materials. In pulse-echo reflection mode, ultrasonic signals reflected by defects possess information about defect size and orientation. In some materials there are many scattering centers such as grain boundaries, which can generate echoes that seem to be randomly distributed in time. The ensemble of these echoes is usually defined as background noise. The noise signals from grain boundaries and other microstructures make a fundamental limit on the detection of small cracks, flaws, or other metallurgical defects. Therefore the signal detection and noise suppression problem is important, particularly for ultrasonic NDE of composite materials and highly scattering materials. Indeed, numerous contributions have attempted to address this problem from different points of view. Such signals have been analyzed in the time domain, frequency domain, or time-frequency/scale domain. Signal averaging, auto- and cross-correlation matched filtering, autoregressive analysis, order statistics, frequency spectrum analysis, split spectrum processing (SSP),[1,2] wavelet transform (WT),[3] and model-based optimization algorithms[4,5] have all been used to analyze ultrasonic signals. In this paper we have no intention of reviewing all these techniques.

In general, if the signal shape expected from a flaw can be estimated, optimal detection can be achieved in the time domain using a matched filter with an impulse response that models the flaw echo. For frequency-domain filtering, it is assumed that the echoes due to the flaws differ in spectral content from those echoes produced by background scattering noise. An optimal filtering approach in the frequency domain is to utilize the Wiener filter. Unfortunately, the Wiener filter requires *a priori* knowledge of the power spectral density characterizations both of the noise and the flaw echoes, which are not available normally in ultrasonic testing. Furthermore, due to the fact that flaw and noise signals usually occupy the same time window, time-domain smoothing or frequency-domain filtering cannot substantially reduce the noise without also reducing the flaw component unless signal-to-noise ratio (SNR) is significantly high. Since the ultrasonic signal is usually a time and frequency limited transient signal in ultrasonic flaw detection, time-frequency analysis techniques are more appropriate to process ultrasonic NDE signals. The most appealing is the SSP technique, which is based on the short time Fourier transform time-frequency decomposition of the ultrasonic signal. In this technique, the ultrasonic signal is divided into a set of narrow band signals using a bank of band-pass filters, to which nonlinear postprocessing is applied to reduce the noise contribution and to extract the flaw information. This technique has been established as an effective method of achieving flaw enhancement and grain noise suppression. The WT is the latest technique of processing transient nonstationary signals in the time-frequency domain. Most recently on the basis of the WT technique a wavelet transform signal processor (WTSP) has been designed for ultrasonic flaw detection, leading to a state-of-the-art denoising performance.[3] In WTSP, the ultrasonic signal is decomposed into the time-

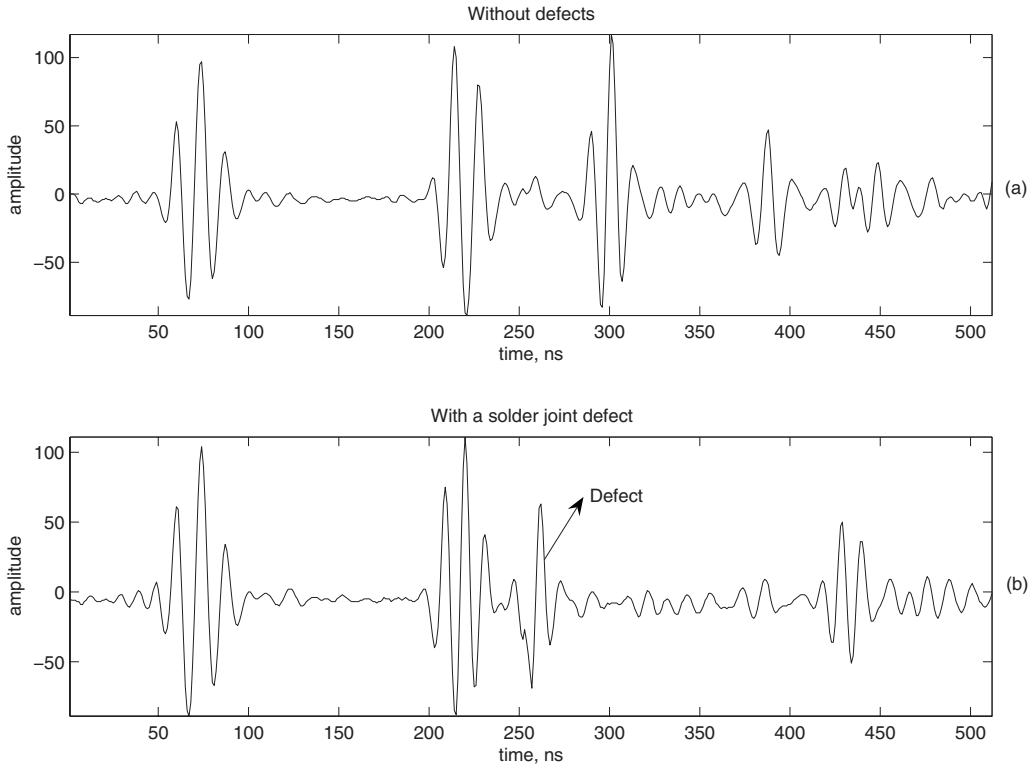


FIG. 1. Typical ultrasonic signals measured from a flip-chip package mounted on a ceramic substrate (a) without defects and (b) with a solder joint defect.

frequency domain by using the continuous wavelet transform (CWT), and *pruning and thresholding* operations are applied to the decomposed coefficients. The processed coefficients are used to reconstruct the signal using the inverse CWT. The pruning procedure replaces the band-pass filtering used in the Fourier-based approaches. The thresholding technique is a near-optimal denoising technique for a wide class of signals corrupted by additive white Gaussian noise.[6–8] Even though this procedure may seem similar to thresholding the signal in the time or frequency domain, the WT thresholding results in minimal averaging or smoothing of the signal in the time or frequency domain because each time-frequency contribution of the original signal is considered locally in the time-frequency domain.

In ultrasonic NDE, the insonified materials, such as composite materials, a homogeneous material with a limited number of defects, biological tissues, microelectronic packages, and aerospace and automobile components, often will be some form of layered structures. Large impedance contrasts typically exist only in a limited number of positions, corresponding to interfaces between different materials or faults. This implies that the reflectivity will be sparse, and the measured ultrasonic signal can be modeled as a sparse signal in these instances.[9,10] The measured ultrasonic signal $y(t)$ acquired from a layered structure in pulse-echo ultrasonic testing can be expressed as a superposition of M-reflected echoes,

$$y(t) = \sum_{i=1}^{M} x_i(t) + \xi(t), \tag{1}$$

where $x_i(t)$ is the ultrasonic echoes of the flaws/internal structures embedded in noise $\xi(t)$. Figure 1 shows typical ultrasonic A-scans measured from a microelectronic package on a commercial scanning acoustic microscopy using a 230 MHz focused transducer.

Sparseness is a powerful constraint that has been exploited extensively in seismic applications. Employing the sparse prior of ultrasonic signals, a new method based on sparse signal representations was designed in this paper for ultrasonic flaw detection and its performance was investigated using numerical results. The sparse signal representation is the most recent technique in the signal processing community, and various algorithms have been proposed.[11–16] The novel technique has found increasing relevance in a large number of application domains, such as signal processing,[12,17] compression,[15] and feature extraction.[18] Recently the sparse representation technique has been utilized to improve the axial resolution of acoustic microimaging by the authors[9,10,19] and ultrasonic echo separation.[20] The potential applications of the sparse representation technique in ultrasonic nondestructive testing have been briefly addressed in our recent paper.[21] In this work, the sparse representation technique is extended to the ultrasonic signal denoising problem, where ultrasonic flaw signals are contaminated by white Gaussian additive noise or correlated noise.

This paper is organized into five sections as follows. In Sec. II, after a brief introduction of the sparse representation model, various sparse representation algorithms are reviewed. The sparse representations of ultrasonic signals are then demonstrated. In Sec. III, pulse detection and noise suppression via sparse and overcomplete representations are described. Numerical results of denoising ultrasonic signals embedded with white Gaussian noise and correlated noise

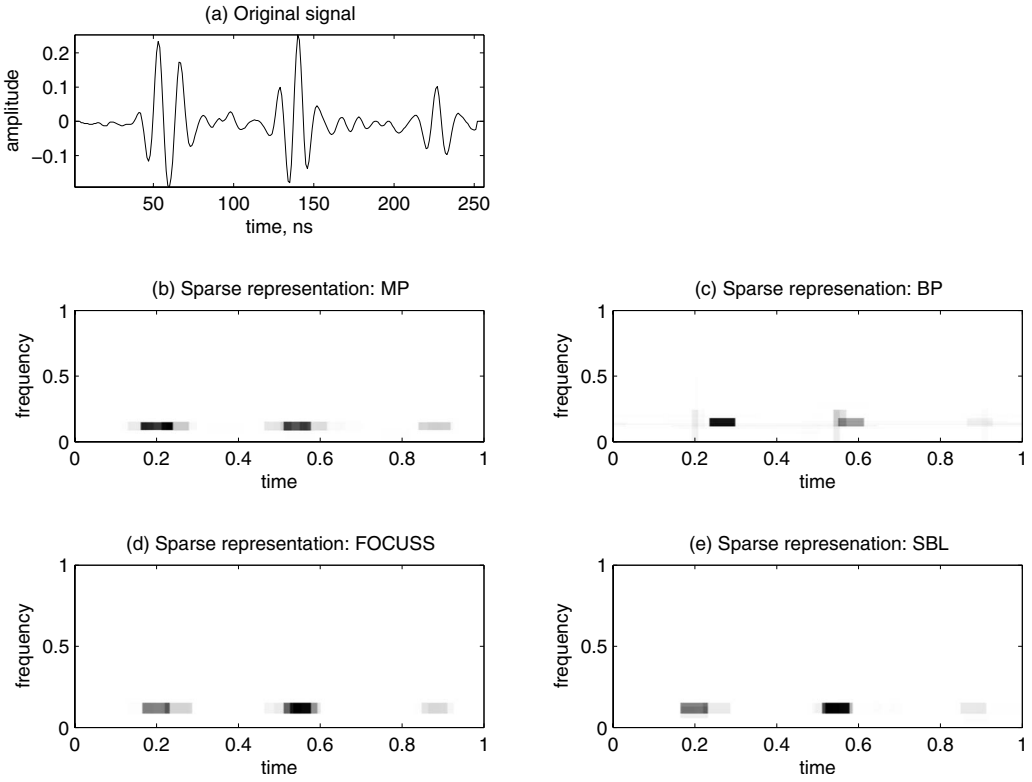


FIG. 2. (a) An ultrasonic signal and its sparse representations obtained by using (b) MP, (c) BP, (d) FOCUSS, and (e) SPL, respectively.

are presented in Sec. IV. Finally, conclusions are given in Sec. V.

II. SPARSE AND OVERCOMPLETE REPRESENTATIONS OF ULTRASONIC SIGNALS

Given a full rank dictionary matrix $\Phi \in R^{N\times L}$ that contains L atoms (or called basis vectors), $\{\phi_i\}_{i=1}^{L}$, as its columns, the representation of a signal is sought as a linear combination of these atoms, i.e.,

$$y = \Phi c, \tag{2}$$

where $c \in R^L$ is the vector of coefficients. Depending on the dictionary dimension, there are three cases: (1) the overdetermined case ($L<N$), (2) the complete case ($L=N$), and (3) the underdetermined case ($L>N$). For the first two cases, the representation may be obtained using principle component analysis[22] and independent component analysis.[23] In the third case, the decomposition is nonunique. The nonuniqueness gives us the possibility of adaptation, i.e., of choosing among many representations one that is most suited to our purposes. The most popular and important representation is the sparse representation, which is of great interest in many applications such as image compression, denoising of signals, and blind source separation. In this paper we focus on the sparse overcomplete representations for denoising of ultrasonic signals.

The sparse overcomplete representation problem is formulated as follows: Given a signal $y \in R^N$ and the overcomplete dictionary Φ, seek the sparsest coefficient vector $c \in R^L$ under the linear model (2). This corresponds to solving the following variational problem:

$$\text{minimize } \|c\|_0 \quad \text{subject to} \quad y = \Phi c \quad (\mathrm{P}_{1a}), \tag{3}$$

where $\|\cdot\|_0$ is the l^0 norm, counting the nonzero entries of a vector. In practical ultrasonic nondestructive detection, ultrasonic signals are commonly contaminated by background noise. The sparse representation of a noisy signal become the following:

$$\text{minimize } \|c\|_0 \quad \text{subject to} \quad y \approx \Phi c \quad (\mathrm{P}_{1b}). \tag{4}$$

Precisely, the rough constraint $y \approx \Phi c$ is replaced with a clear requirement to allow a bounded representation error, $\|y - \Phi c\|_2 \leq \varepsilon$.

Directly solving (P_{1a}) proves to be an nondeterministic polynomial time (NP)-hard problem. Hence, approximate solutions are usually explored instead. In the past decade or so several efficient approaches have been developed. The simplest one is the matching pursuit (MP) algorithm,[11] which is a greedy algorithm that selects the dictionary atoms sequentially. Figure 2(a) plots an ultrasonic signal extracted from the measured A-scan shown in Fig. 1(a) and Fig. 2(b) displays in time-frequency domain the sparse representation of Fig. 2(a) obtained using the MP algorithm. The darkness of the time-frequency image increases with the energy value, and each time-frequency atom selected by the MP method is represented by a Heisenberg box.[11]

A second well-known sparse basis selection approach is the basis pursuit (BP) algorithm.[12] Under certain conditions, which have been found empirically and theoretically,[12,24] problem (P_1) can be considered equivalent to the convex optimization problem given by the following:

$$\text{minimize } \|c\|_1 \quad \text{subject to} \quad y \approx \Phi c \quad (P_2). \tag{5}$$

Another sparse basis selection approach is the focal underdetermined system solver[13] (FOCUSS), which approximates the problem (P_1) using the l^p norm with $p \leqslant 1$ as a replacement to the l^0 norm, i.e.,

$$\text{minimize } \|c\|_p \quad \text{subject to} \quad y \approx \Phi c \quad (P_3). \tag{6}$$

Here, for $p<1$ the similarity to the true sparsity measure is better, but the overall problem becomes nonconvex, giving rise to local minima that may divert the optimization. The BP and FOCUSS essentially use a fixed regularizing weight prior of the form

$$p(c) \sim \exp\left(-\sum_{i=1}^{L} |c_i|^p\right), \tag{7}$$

where $p \in (0,1]$ to encourage sparsity. More recently the sparse Bayesian learning (SBL) has been adapted to the problem of finding sparse representations of signals from overcomplete dictionaries.[14] In contrast to FOCUSS and BP, which assume a fixed prior, SBL estimates a flexibly parametrized prior from the observed data. The parametric form of the prior is given by a multivariate Gaussian distribution, i.e.,

$$p(c) = \prod_{i=1}^{L} (2\pi\alpha_i)^{-1/2} \exp(-c_i^2/2\alpha_i), \tag{8}$$

where $\alpha=[\alpha_1, \ldots, \alpha_L]^T$ is a vector of L hyperparameters controlling the prior variance of each c_i. These hyperparameters are then estimated from the observed data y. The sparse representations of Fig. 2(a) obtained using BP, FOCUSS, and SBL are presented in Figs. 2(c)–2(e), respectively. Here, a customized dictionary described in Sec. III was used to perform the decomposition for MP and FOCUSS and SBL while a wavelet package dictionary with the mother wavelet "Symmlet 10" (Ref. 25) was used to perform the BP decomposition due to computational complexity. From the demonstration of Fig. 2, it can be seen that the ultrasonic flaw echoes have been decomposed into a few significant coefficients by these sparse representation approaches.

III. PULSE DETECTION AND NOISE SUPPRESSION USING SPARSE BAYESIAN LEARNING

Recent years have witnessed a growing interest in sparse signal representations. Using redundant representations and sparsity as driving forces for image denoising has drawn a lot of research attention in the past decade or so.[25–28] The introduction of sparse signal representations gave rise to the ability to address the signal denoising problem as a direct sparse decomposition technique over overcomplete dictionaries. Although under an overcomplete dictionary the decomposition of a signal is underdetermined, recent research shows that this can offer great advantages in many applications, for example, image restoration and data compression. One is that there is greater flexibility in capturing structure in the data.[15] Instead of a small set of general basis vectors, there is a larger set of more specialized atoms such that relatively few are required to represent any particular signal. These can form more compact representations because each atom can describe a significant amount of structure in the data. The second is super-resolution.[12] We can obtain a resolution of sparse objects that is much higher than that possible with traditional methods. An additional advantage is that overcomplete representations increase stability of the representation in response to small perturbations of the signal. In addition, the redundant representations have the desired shift invariance property.[25]

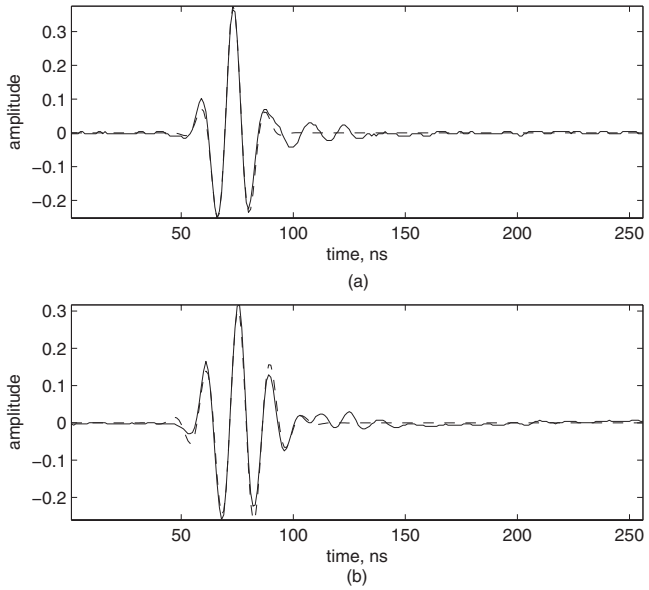


FIG. 3. The Gabor approximation (dashed line) of normalized ultrasonic echoes (solid line) acquired using a 230 MHz focused transducer with (a) low excitation and (b) high excitation.

Pulse detection consists of determining the presence or absence of a pulse and estimating its amplitude and arrival time. Noise suppression is to obtain a signal $\hat{y}(t)$ as close as possible to $\Sigma_{i=1}^{M} x_i(t)$ in the noisy ultrasonic signal model (1). These advantages of sparse overcomplete representations will benefit pulse detection and noise suppression in ultrasonic testing. For ultrasonic signals, ultrasonic flaw echoes are the significant structures. If the dictionary atoms are designed to match the ultrasonic flaw echoes well, $\Sigma_{i=1}^{M} x_i(t)$ will be packed into very few significant coefficients while the noise energy will spread out around the whole time-frequency domain due to its mismatch with dictionary atoms. This will greatly facilitate the application of denoising methods, for instance, wavelet thresholding.

Obviously, in order to obtain compact representations of a signal, designing an appropriate dictionary is important. The choice of the dictionary for a particular problem improves the signal processing capability of the sparse presentation technique. Tailoring a dictionary to the actual problem is possible and should be done. The ultrasonic signal is usually a broadband pulse modulated at the center frequency of the transducer and is usually modeled as a Gabor function.[29] In order to examine the efficiency of the Gabor model for approximating ultrasonic pulses in our system, ultrasonic echoes reflected from the front surface of a steel block in a water tank were collected using a 230 MHz focused transducer. Two collected echoes are normalized and shown in Fig. 3. The one in Fig. 3(a) was obtained with low energy

transducer excitation while the one in Fig. 3(b) was with high power. The measured echoes were then fitted by the cosine Gabor model. The root-mean-squared error of disparity between the estimated and measured echoes was small at 1.12% and 1.11%, respectively. Therefore, an overcomplete multiscale Gabor dictionary was customized for our application. The dictionary consisted of several fixed scale critically sampled cosine Gabor bases. Each atom was defined by the parameters $\gamma=(s,u,v)$ as

$$g_\gamma = (A/\sqrt{s})\exp^{-\pi(t-u)^2/s^2}\cos(v(t-u)), \tag{9}$$

where s is the scale of the function, $s \in \{lms:ds:rms\}$; u its translation, $u \in \{0:du:N-1\}$; v its frequency modulation, $v \in \{lmv:dv:rmv\}$; and constant and factor $A/\sqrt{s}$ normalize g_γ. The parameters $\gamma=(s,u,v)$ were discretized as follows: $ds=s_r/10$, $du=2$, $dv=v_r/10$, $lms=6ds$, $rms=s_r$, $lmv=6dv$, and $rmv=v_r$, where the reference parameters $\gamma_r=(s_r,u_r,v_r)$ were obtained by approximating the reference ultrasonic echo shown in Fig. 3(a) using the Gabor model. In Fig. 3(a), $s_r=25.6289$ and $v_r=0.3994$. The sampling scheme is motivated as follows: (1) Suppose the pulse variation is slow, and the ultrasonic echoes in ultrasonic signals should be well fitted by an atom with parameters close to γ_r, (2) the echoes reflected inside the package have lower frequencies than the reference echo due to frequency dependent attenuation, (3) control the dictionary size to save computation time and memory, and (4) the dictionary design including the dictionary size is very important as it determines the matching degree between the dictionary atoms and ultrasonic echoes, and the sparsity condition under which the solution of (P_2) and (P_3) is also the solution of (P_1).[24] The dictionary used in this paper was determined empirically. Notice that the used sampling restrictions are not necessary for the proposed technique. They are only used to customize the dictionary for this particular application. See Ref. 30 for details and other sampling schemes.

Once we have defined a dictionary, a sparse representation approach needs to be chosen to obtain the sparse representation of an ultrasonic signal. Because of the high nonlinear approximation rate, the adaptive MP decomposition provides an effective denoising scheme by only selecting the significant echoes. The significant echoes are atoms for which $|\langle R^k y, \phi_k\rangle|/\|R^k y\| >$ noise, where $R^k y$ is the residue after the kth iteration.[11] MP was applied for ultrasonic flaw detection with moderate to good success. However, when the dictionary is not orthogonal, MP may fail to super-resolve and not be sparsity preserving, especially when the SNR is very low or two echoes are closely spaced. This failure is due to the greediness of the algorithm, which may lead MP to wander off in the wrong direction of many atoms. The BP method was developed to overcome the problem, and good results have been reported.[12] BP, on the other hand, suffers from the fact that the l^1 prior distribution of coefficients c is hardly to be true for our ultrasonic signals. The l^1-norm has a tendency to equally distribute the energy amongst atoms with similar correlations. The l^p-like prior distribution of c in the FOCUSS method, where p can be set as any value between 0 and 1, provides the flexibility to set a prior close to a particular signal. Here a further issue arises: how to accurately set the prior parameter p in FOCUSS? In fact, the exact value of p is usually unknown in ultrasonic NDE. For some applications such as layered structures, a reasonable good estimation of p might be made in advance by using the prior information of samples. Another alternative is the SBL approach, where the prior parameter p is estimated from the observed data, avoiding the need to set a fixed value to it. The performance of SBL, FOCUSS, and BP has been compared by examining the relative frequency of cases where each algorithm failed to uncover the generating sparse representations c by a simulation study in Ref. 14 There, Wipf and Rao showed that SBL retains a desirable property of the l^0-norm (i.e., no structural errors as occur with BP) while often possessing a more limited constellation of local minima (i.e., fewer convergence errors than with FOCUSS), where convergence error means convergence to a standard local minimum, and structural error means convergence to a minimum (possibly global) that is not maximally sparse yet and has a lower cost function value than the generating solution.[14] In the light of the above analysis, the SBL approach has been chosen to perform the sparse decomposition in this work.

Pulse detection and noise suppression is carried out on the decomposed coefficients c in the time-frequency domain. Similar procedures used in WTSP, including pruning and thresholding, are employed in the proposed technique to achieve the goal. Because the significant ultrasonic echoes $x_i(t)$ are band limited, the corresponding atoms of significant coefficients contributed by $x_i(t)$ are limited to a range of frequency band with $v \in (v_1,v_2)$, where v is the center-frequency parameter of a dictionary atom in Eq. (9). The pruning operation is expressed mathematically as

$$c_i = \begin{cases} 0 & \text{for } v < v_1 \\ c_i & \text{for } v_1 < v < v_2 \\ 0 & \text{for } v > v_2. \end{cases} \tag{10}$$

The pruning operation would be equivalent to band-pass filtering used in Fourier-based approaches, resulting in the elimination of coefficients outside the range (v_1,v_2), which are most likely to be from noise. The difference is that sparse representations are high compact representations, so the pruning operation is more efficient than band-pass filtering in the frequency domain.

The thresholding operation is a celebrated denoising method in wavelet based signal processing, which utilizes the sparsity of the unitary wavelet coefficients and achieves a near-optimal performance for a wide class of signals corrupted by additive white Gaussian noise. The thresholding operation is expressed as

$$\hat{c}_i = \begin{cases} 0 & \text{for } |c_i| < \lambda \\ \mathrm{sgn}(c_i)\cdot(|c_i| - \lambda) & \text{for } |c_i| \geqslant \lambda. \end{cases} \tag{11}$$

Because of the high compact essence of sparse representations, the significant ultrasonic echoes are packed into a few significant coefficients, and noise energy is scattered all over the dictionary atoms, generating insignificant coefficients. Hence, the thresholding procedure could be even more efficient than in the WT domain. Notice that this approach is

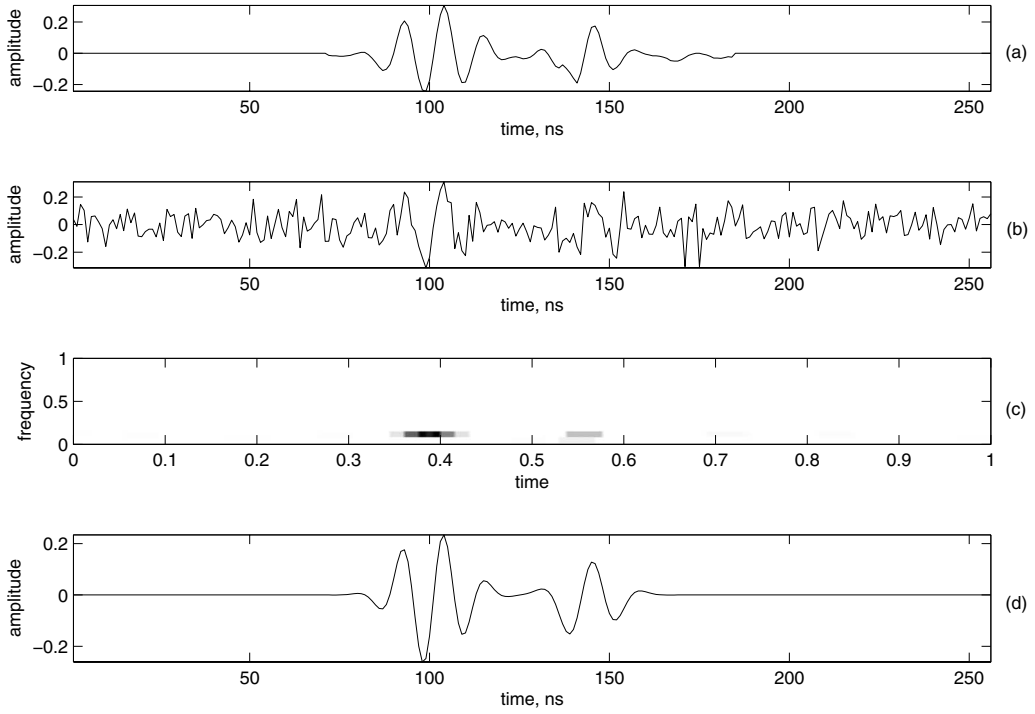


FIG. 4. Pulse detection and noise suppression using the proposed technique. (a) Original ultrasonic signal (b) with white Gaussian noise added. (c) The sparse representation of the resulting noisy signal. (d) The noise reduction output.

optimal for white noise, but it can fail if the noise is correlated. Fortunately, unlike the WT where daughter wavelets are generated from the mother wavelet by dilation and shift operations, in the proposed technique the dictionary atoms could be designed to match the significant ultrasonic echoes maximally and mismatch the correlated noise, relieving the impact of correlated noise on pruning and thresholding procedures. For example, a learned dictionary[16,31] could be employed to achieve this requirement.

In many ultrasonic applications, the only information available is the center frequency of the transducer used and its frequency bandwidth, while very little is known about the noise. For this reason, the center frequency and the bandwidth in combination with the consideration of the frequency-dependent attenuation of materials and the structures of the used samples are thus utilized to determine the band range (v_1, v_2) in the pruning operation. The threshold λ is determined empirically, which can be set as a function of noise level. Both (v_1, v_2) and λ are kept constant for a particular application.

The final step of the proposed technique is to reconstruct the ultrasonic signal using the processed coefficients $\hat{c}_i$ and the corresponding dictionary atoms using Eq. (2). Notice that in WTSP the inverse CWT defined as

$$\hat{y}(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_y(a,b) h_{a,b}(t) \frac{da}{a^2} db \tag{12}$$

is not a perfect reconstruction algorithm, where a is the dilation, b is the shift, $h_{a,b}(t)$ are the daughter wavelets, and $W_y(a,b)$ are the processed wavelet coefficients. The amplitudes of reconstructed echoes are thus inaccurate. For sparse representation approaches, the signal reconstruction is perfect, and the amplitudes of reconstructed echoes are more accurate. This advantage of the proposed technique is important in ultrasonic NDE for defect sizing.

IV. EXPERIMENTAL RESULTS

In ultrasonic flaw detection, the probability of detection usually decreases as the defect size decreases while the probability of false indication increases. This is mainly due to the effect of background noise masking the defect signal. Denoising of ultrasonic signals is thus an extremely important issue in ultrasonic NDE in order to correctly identify and measure smaller defects. The following experiments are designed to examine the performance of the proposed approach for pulse detection and noise suppression.

A. Noise reduction for signals with additive white Gaussian noise

This experiment is utilized to verify the signal detection improvement for ultrasonic signals embodied in white noise. An experimental result obtained using the proposed technique is presented in Fig. 4, where the original ultrasonic signal $y(t)$ is plotted in Fig. 4(a). Computer generated white Gaussian noise $\xi(t)$ was added to the waveform, and the resulting noisy signal was processed using the proposed technique. A noisy signal with SNR=−2.82 dB is displayed in Fig. 4(b), and its sparse representation decomposed by SBL is presented in Fig. 4(c). The processed output is shown in Fig. 4(d). It is observed that the ultrasonic echoes have clearly been detected and recovered from the noisy signal.

For the purpose of comparison, Fig. 5 presents the noise reduction result obtained using WTSP, where the Morlet wavelet has been used. Figure 5(b) shows the time-frequency representation of the same noisy signal of Fig. 4(b), which

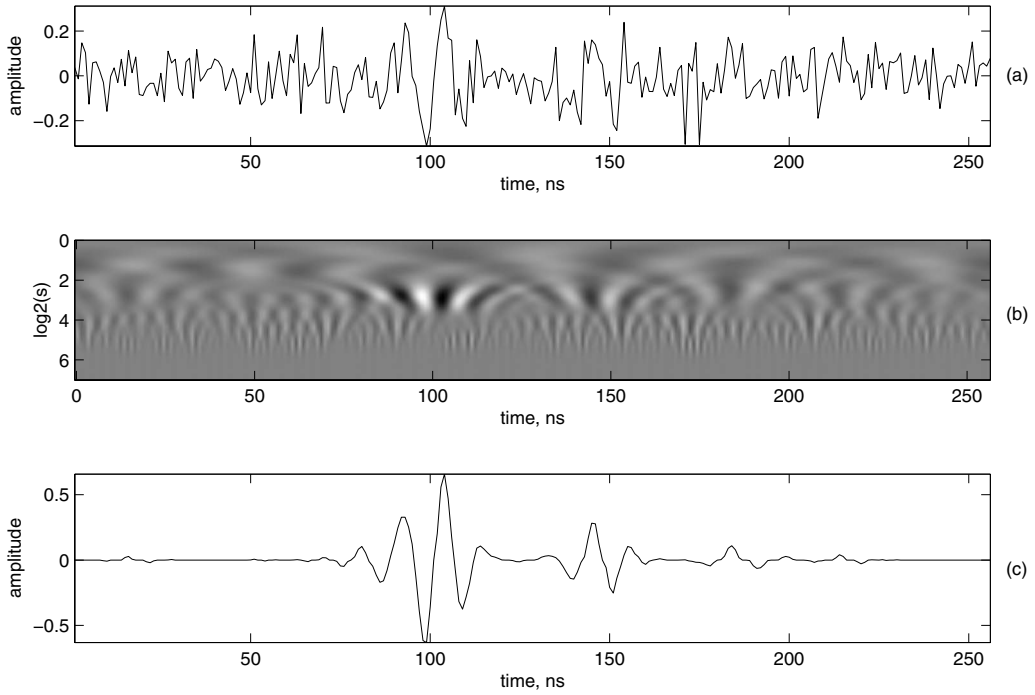


FIG. 5. Pulse detection and noise suppression using the WTSP approach. (a) The noisy signal as shown in Fig. 4(b). (b) The time-frequency representation obtained by the CWT. (c) The WTSP output.

was obtained by applying the CWT algorithm. The WTSP output is displayed in Fig. 5(c). Comparing Figs. 4 and 5, it can be seen that the proposed technique outperforms the WTSP approach

In order to quantify the efficiency of the proposed approach in ultrasonic signal detection, a similar experiment to Abbate *et al.*[3] has been performed in this work. The output SNR was evaluated as a function of the input SNR. The output SNR_{out} is calculated as the ratio of the output signal power computed in a time window where the ultrasonic echoes are present and another time window for which the original echoes are null. SNR_{in} is calculated as the ratio of the power of $y(t)$ and the power of $\xi(t)$ added to it. Figure 6 presents our experimental results. From Fig. 6, it can be seen that when SNR_{in} is not very low, the proposed approach is superior to WTSP. However, when SNR_{in} is extremely low, decomposing the noisy signal is not reliable. Although the two flaw echoes might still be detected, some noise might also be detected wrongly as flaw echoes. When $SNR < -5$ dB in our experiments, the small echo in Fig. 4(a) might be lost. The unreliable results are most likely caused by the following: One is the convergence errors existing in sparse decomposition using SBL. When SNR_{in} decreases the convergence errors increase. Indeed, this has been observed by Wipf and Rao.[14] Another reason is that in SBL, the noise level is estimated from the observed data inside the SBL algorithm, and very low SNR likely leads to a big estimation error.

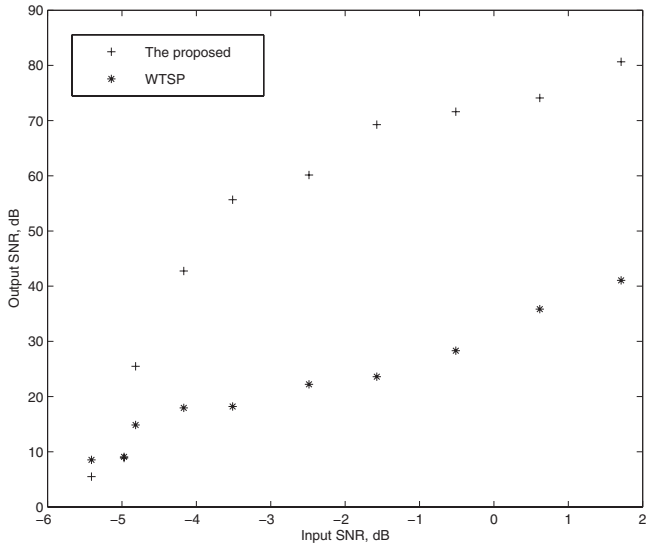


FIG. 6. Output SNR against the input SNR for the WTSP approach and the proposed sparse representation based approach.

There are two possible remedies that could alleviate the problem. One is to use the BP denoising algorithm[12] to replace SBL because there is no convergence error for BP. The dictionary design including the dictionary size is, however, vital for BP to avoid a big structural error as it determines the sparsity condition under which the solution of (P_2) is also the solution of (P_1).[24] Second is to learn a dictionary to maximally match the dictionary atoms and the ultrasonic echoes while mismatching the noise. These issues will be investigated in the future.

B. Noise reduction for signals with correlated noise

In ultrasonic NDE the background noise arising from nonflaw related scattering of the reflected incident ultrasonic wave can be generally modeled as white Gaussian noise in most applications. When inspecting highly scattering materials and imaging biological tissues, however, grain boundary

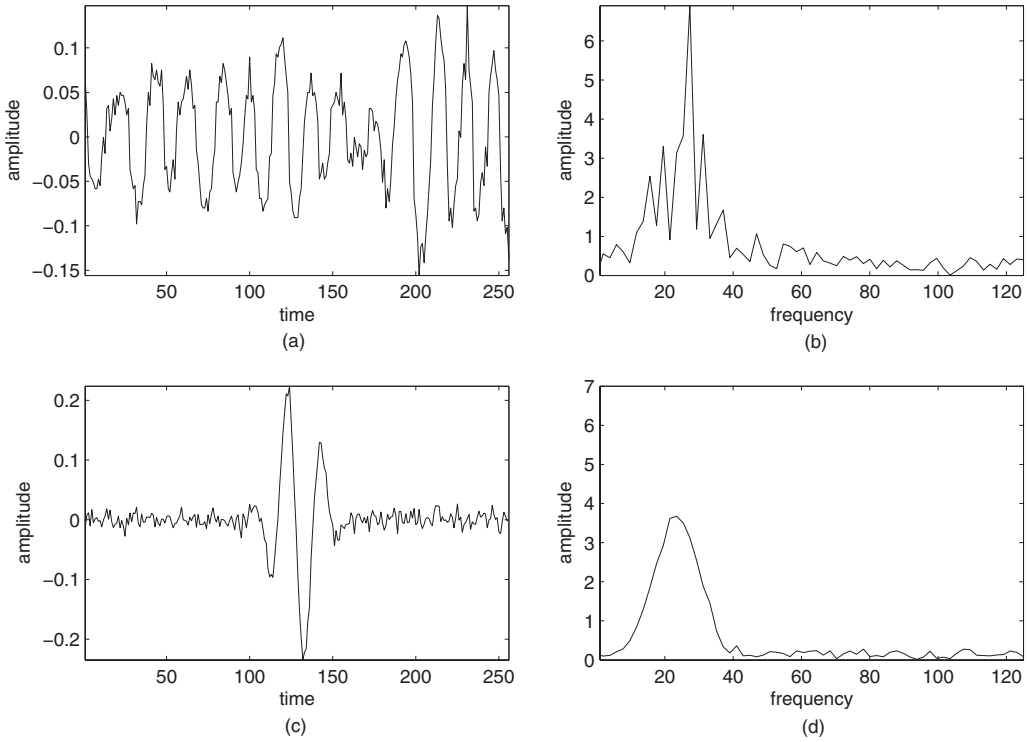


FIG. 7. (a) Correlated noise obtained experimentally. (b) Power spectrum of the noise. (c) An ultrasonic flaw signal. (d) Power spectrum of the flaw signal.

scattering will exhibit a small-scale correlation. The grain noise is thus modeled as correlated noise. The ultrasonic flaw detection problem is then summarized as separating the contribution of a coherent scatterer from unresolvable scatterers such as grain noise.

Conventional filtering techniques, which utilize whitening of the noise with subsequent matched filtering, do not perform satisfactorily improving the SNR of transient wideband signals in correlated noise, largely due to the fact that the signal and noise frequency spectra overlap in the frequency domain and are not easily separable.[3] These methods utilize any spectral difference between the flaw echo and the noise to filter out the noise contributions and enhance the frequencies of interest.

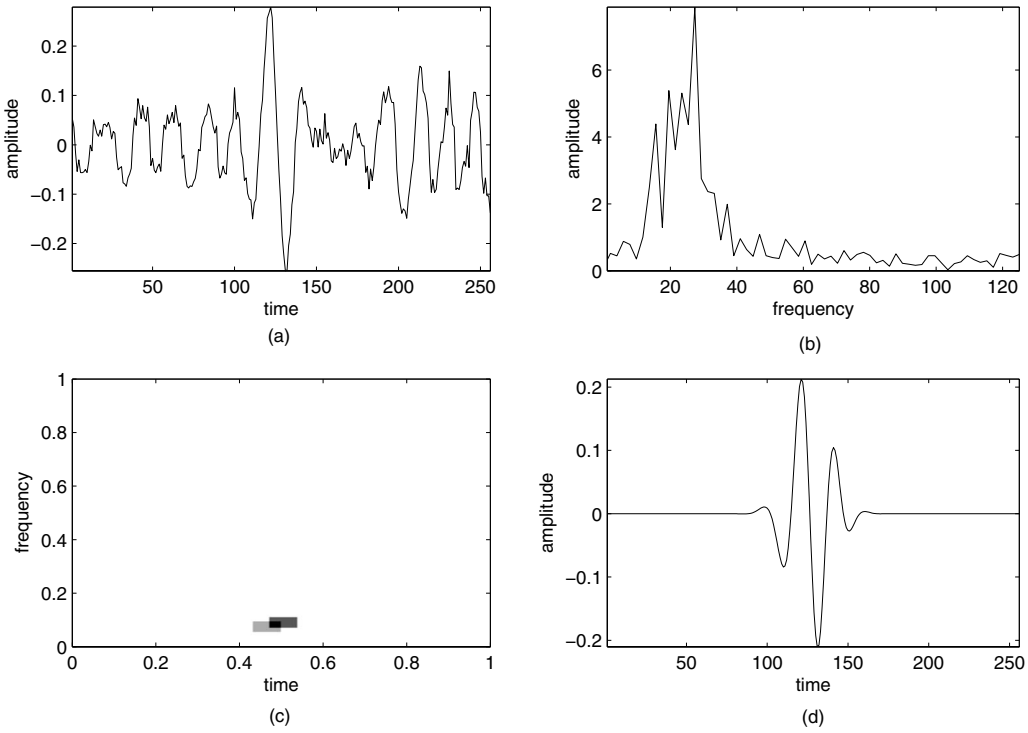


FIG. 8. (a) The noisy ultrasonic signal obtained by adding the correlated noise in Fig. 7(a) to the flaw signal in Fig. 7(c). (b) Its power spectrum. (c) The sparse representation of (a). (d) The noise reduction output using the proposed technique.

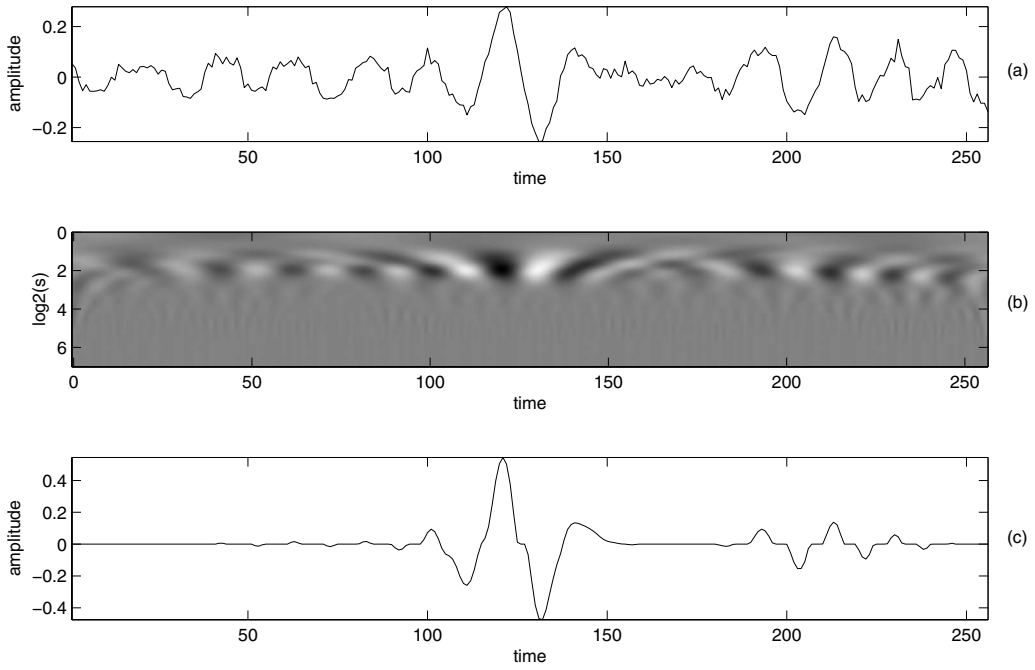


FIG. 9. (a) The noisy signal shown in Fig. 8(a). (b) The time-frequency representation obtained by the CWT. (c) The WTSP output.

This experiment is carried out to examine the efficiency of the proposed technique for noise suppression of ultrasonic signals with correlated noise. A noise acquired from an immersion sample using a 30 MHz focused transducer is shown in Fig. 7(a), and its power spectrum is shown in Fig. 7(b). An ultrasonic flaw signal acquired using the same transducer is shown in Fig. 7(c), and its power spectrum is shown in Fig. 7(d). An artificial noisy signal as shown in Fig. 8(a) was created by adding the noise in Fig. 7(a) to the flaw signal in Fig. 7(c), and its power spectrum is shown in Fig. 8(b). The corresponding SNR of the noisy signal was −1.85 dB. From Fig. 7 it can be seen that the signal and noise spectra are severely overlapped. The sparse representation of the noisy signal decomposed by SBL is shown in Fig. 8(c). The processed output is presented in Fig. 8(d). The corresponding WTSP processing result of the noisy signal is presented in Fig. 9 for comparison. It is clear that the sparse representation based technique outperforms the popular WTSP approach not only in noise suppression but also in reconstruction of ultrasonic echo amplitude and shape.

C. Position error

In ultrasonic NDE, we often need to estimate the pulse arrival time. This is used for defect location, thickness measurement, and sound velocity determination. An experiment similar to Sec. IV A was carried out to verify the positioning performance of the proposed technique, which was evaluated by measuring the change in time delay between two pulses. The time delay was measured in terms of sample number and in absolute values. An original signal different from Fig. 4(a) was used so that the two pulses are further away. In fact, the original time delay is 148 sample numbers in the experiment. The amplitudes of the two pulses are comparable. The error of measured time delay is presented in Table I as a function of SNR_{in}. Notice that when we customized the Gabor dictionary in Sec. III, the parameter $du=2$ so that the measured error of time delay is multiple of 2. The results obtained using WTSP are presented as well in the table for comparison. It is seen that both approaches measured the time delay accurately within two samples for SNR down to −5.79 dB.

V. CONCLUSIONS

A sparse and redundant representation based ultrasonic signal processing approach has been proposed for improving ultrasonic flaw detection. In this work, an ultrasonic signal is decomposed into sparse representations over a customized overcomplete Gabor dictionary using a sparse signal representation algorithm, where the ultrasonic flaw echoes are packed into a few significant coefficients while the noise energy tends to spread along the whole dictionary with small coefficients. This facilitates the separation of noise and flaw

TABLE I. The error of measure delay time between two pulses against the SNR. Results are presented in sample number.

		SNR_{in} (dB)										
		15.62	6.97	1.54	0.22	−0.74	−1.61	−2.58	−3.23	−4.14	−4.72	−5.79
Error of time delay (in sample number)	The proposed	0	0	0	0	0	0	2	0	2	2	2
	WTSP	0	0	0	0	1	1	1	1	1	1	2

signals. Ultrasonic flaw detection and noise suppression is then carried out on the basis of the obtained sparse representations by adapting the WT thresholding technique into the proposed approach. The proposed approach has been compared with the state-of-the-art WTSP. Experimental results have demonstrated its superior performance.

[1]X. Li, N. M. Bilgutary, and R. Murthy, "Spectral histogram using the minimization algorithm: Theory and applications to flaw detection," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 279–284 (1992).

[2]V. Newhouse, N. Bilgutay, J. Saniie, and E. Furgason, "Flaw to gain echo enhancement by split spectrum processing," Ultrasonics **20**, 59–68 (1982).

[3]A. Abbate, J. Koay, J. Frankel, S. Schroeder, and P. Das, "Signal detection and noise suppression using a wavelet transform signal processor: Application to ultrasonic flaw detection," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 14–26 (1997).

[4]L. Satyanarayan, K. Bharath Kumaran, C. V. Krishnamurthy, and K. Balasubramaniam, "Inverse method for detection and sizing of cracks in thin sections using a hybrid genetic algorithm based signal parameterisation," Theor. Appl. Fract. Mech. **49**, 185–198 (2008).

[5]R. Demirli and J. Saniie, "Model-based estimation of ultrasonic echoes Part I: Analysis and algorithms," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 787–802 (2001).

[6]D. L. Donoho, "De-noising by soft thresholding," IEEE Trans. Inf. Theory **41**, 613–627 (1995).

[7]D. L. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," Ann. Stat. **26**, 879–921 (1998).

[8]D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," Biometrika **81**, 425–455 (1994).

[9]G. M. Zhang, D. M. Harvey, and D. R. Braden, "An improved acoustic microimaging technique with learning overcomplete representation," J. Acoust. Soc. Am. **118**, 3706–3720 (2005).

[10]G. M. Zhang, D. M. Harvey, and D. R. Braden, "Adaptive sparse representations of ultrasonic signals for acoustic microimaging," J. Acoust. Soc. Am. **120**, 862–869 (2006).

[11]S. G. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," IEEE Trans. Signal Process. **41**, 3397–3415 (1993).

[12]S. S. Chen, "Basis pursuit," Ph.D. thesis, Stanford University, 1995.

[13]B. D. Rao and K. Kreurz-Delgado, "An affine scaling methodology for the best basis selection," IEEE Trans. Signal Process. **47**, 187–200 (1999).

[14]D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," IEEE Trans. Signal Process. **52**, 2153–2164 (2004).

[15]M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," Neural Comput. **12**, 337–365 (2000).

[16]M. Aharon, M. Elad, and A. M. Bruckstein, "K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," IEEE Trans. Signal Process. **54**, 4311–4322 (2006).

[17]M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Trans. Image Process. **15**, 3736–3745 (2006).

[18]O. Williams, A. Blake, and R. Cipolla, "Sparse Bayesian learning for efficient visual tracking," IEEE Trans. Pattern Anal. Mach. Intell. **27**, 1292–1304 (2005).

[19]G. M. Zhang, D. M. Harvey, and D. R. Braden, "Advanced acoustic microimaging using sparse signal representation for the evaluation of microelectronic packages," IEEE Trans. Adv. Packag. **29**, 271–283 (2006).

[20]G. M. Zhang, D. M. Harvey, and D. R. Braden, "Effect of sparse basis selection on ultrasonic signal representation," Ultrasonics **45**, 82–91 (2006).

[21]G. M. Zhang, D. M. Harvey, and D. R. Braden, "Microelectronic package characterisation using scanning acoustic microscopy," NDT & E Int. **40**, 609–617 (2007).

[22]I. T. Jolliffie, *Principle Component Analysis* (Springer-Verlag, New York, 1986).

[23]S. Roberts and R. Everson, *Independent Component Analysis: Principles and Practice* (Cambridge University Press, Cambridge, UK, 2001).

[24]D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization," Proc. Natl. Acad. Sci. U.S.A. **100**, 2197–2202 (2003).

[25]R. Coifman and D. L. Donoho, "Translation invariant de-noising," *Wavelets and Statistics*, Lecture Notes in Statistics (Springer-Verlag, New York, 1995), pp. 125–150.

[26]W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," IEEE Trans. Pattern Anal. Mach. Intell. **13**, 891–906 (1991).

[27]J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," IEEE Trans. Image Process. **11**, 670–684 (2002).

[28]R. Eslami and H. Radha, "Translation-invariant contourlet transform and its application to image denoising," IEEE Trans. Image Process. **15**, 3362–3374 (2006).

[29]G. Zhang, C. Hou, Y. Wang, and S. Zhang, "Optimal frequency-to-bandwidth ratio of wavelet in ultrasonic non-destructive evaluation," Ultrasonics **39**, 13–17 (2001).

[30]*Gabor Analysis and Algorithms: Theory and Application*, edited by H. G. Feichtinger and T. Strohmer (Birkhauser, Cambridge, MA, 1998).

[31]K. Kreurz-Delgado, J. F. Murray, and B. D. Rao, "Dictionary learning algorithms for sparse representation," Neural Comput. **15**, 349–396 (2003).

Joint deconvolution and classification with applications to passive acoustic underwater multipath

Hyrum S. Anderson[a] and Maya R. Gupta[b]
Department of Electrical and Computer Engineering, University of Washington, Seattle, Washington 98195

(Received 30 April 2008; revised 5 August 2008; accepted 6 August 2008)

This paper addresses the problem of classifying signals that have been corrupted by noise and unknown linear time-invariant (LTI) filtering such as multipath, given labeled uncorrupted training signals. A *maximum a posteriori* approach to the deconvolution and classification is considered, which produces estimates of the desired signal, the unknown channel, and the class label. For cases in which only a class label is needed, the classification accuracy can be improved by not committing to an estimate of the channel or signal. A variant of the quadratic discriminant analysis (QDA) classifier is proposed that probabilistically accounts for the unknown LTI filtering, and which avoids deconvolution. The proposed QDA classifier can work either directly on the signal or on features whose transformation by LTI filtering can be analyzed; as an example a classifier for subband-power features is derived. Results on simulated data and real Bowhead whale vocalizations show that jointly considering deconvolution with classification can dramatically improve classification performance over traditional methods over a range of signal-to-noise ratios.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2981046]

PACS number(s): 43.60.Pt, 43.60.Lq, 43.60.Bf, 43.30.Sf [EJS] Pages: 2973–2983

I. INTRODUCTION

Many signal processing applications require classifying a signal $z(t)$ that has been corrupted by an unknown linear time-invariant (LTI) filter $h(t)$,

$$z(t) = h(t) * x(t) + w(t), \tag{1}$$

where $x(t)$ is the signal of interest, $w(t)$ is additive noise, and $*$ denotes convolution. For example, the LTI filter $h(t)$ could model seismic reflections of an impulsive source, blurring of celestial bodies observed through the Earth's atmosphere, or the effects of an underwater channel on sound propagation. We assume that n labeled training pairs $\{x_i(t), y_i\}$, $i = 1, \ldots, n$, are available to classify the observed signal $z(t)$. As is standard in classifier theory, we assume each $x_i(t)$ and its class label y_i are drawn independently and identically (i.i.d.) from the same joint distribution as the test signal $x(t)$ and its unknown label y. This paper presents joint deconvolution and classification methods, in which the existence of training data can inform an otherwise blind-deconvolution problem. From a classification perspective, the challenge is to deal with the mismatch between training pairs $\{x_i(t), y_i\}$ provided in signal space ("$x(t)$ space" or x-space) and the observed signal $z(t)$ in measurement space (z-space).

The framework developed in this paper will apply to any random LTI filtering, but our emphasis will be on multipath, which can be modeled by an impulse response with sparse coefficients that generally decay with time. Multipath affects many sensing modalities (e.g., ultrasound, radar, terahertz imaging); in this work we present experiments for classifying passive acoustic signals corrupted by multipath in a shallow ocean channel using a single hydrophone at low signal-to-noise ratios (SNRs). For the passive sonar problem, $z(t)$ represents the in-channel received signal, $h(t)$ represents the multipath, and $x(t)$ is the free-field signal. Underwater multipath channels are generally time-varying, that is, the multipath $h(t)$ in Eq. (1) changes between successive transmissions, but not during the transmission. Multipath is highly sensitive to the locations of the source and receiver, making it difficult to model effectively.[1,2] In this paper, we account for the uncertainty in the channel by treating $h(t)$ as a random process.

The main contributions of this paper include: (1) a unified *maximum a posteriori* (MAP) framework for deconvolution and classification for multipath, (2) a quadratic discriminant analysis (QDA) classifier that probabilistically takes into account unknown LTI filtering, (3) and a comparison of feature-based classifiers for marine mammal identification using real whale vocalizations in an acoustically accurate multipath environment.

First, we review related research in Sec. I A. Then in Sec. II, we unify deconvolution and classification in a joint MAP framework. This method jointly estimates a clean signal $\hat{x}(t)$, a channel estimate $\hat{h}(t)$ and a class label y^*. In Sec. III, we argue that if signal estimate $\hat{x}(t)$ is not needed, better classification performance can be achieved by not committing to a particular signal or channel estimate. We show how a QDA classifier can be designed to incorporate the effects of uncertain $h(t)$. The joint MAP deconvolution/classification (joint MAP) and joint QDA deconvolution/classification (joint QDA) methods presented in Secs. II and III classify $z(t)$ based on the entire time signal. We show in Sec. IV how to extend the joint QDA classifier for use with subband power features of $z(t)$. We demonstrate the importance of taking into account the second-order statistics using simulated multipath with mock signals and with real marine

[a]Electronic mail: hyrum@ee.washington.edu
[b]Electronic mail: gupta@ee.washington.edu

mammal vocalizations. We conclude in Sec. V with a discussion about the techniques presented and suggest directions for future research.

A. Related work on classifying signals corrupted by unknown LTI filtering

Signal processing researchers in underwater passive acoustics have considered the problem of classifying signals corrupted by multipath for over 30 years.[3] Ehrenberg *et al.* demonstrated in an ocean acoustic propagation experiment that multipath effects generally cannot be ignored, and that simple time-gating of the received signal can discard too much of the signal information for classification.[4,5] Multipath induced by a shallow ocean channel presents an additional challenge in that the multipath propagation is generally time varying and the structure of $h(t)$ is sensitive to spatial location, making it difficult to estimate or model effectively.[1,2]

A review of the literature reveals four general strategies for classifying signals corrupted with multipath. The first is to extract features from training signals $\{x_i(t)\}$ and the received signal $z(t)$ that are invariant to multipath distortion, then classify based on the multipath-invariant features. Shin *et al.* consider a number of time-frequency features for clutter rejection.[6] Strausberger *et al.* compare different distance measures for 1-nearest-neighbor (1-NN) classification of signals passed through Rician channels for over-the-horizon radar.[7] Recently, Okopal and Loughlin developed features invariant to channel dispersion and dissipation and demonstrated superior classification performance compared to temporal and spectral moment features.[8] In general, classification using channel invariant features can provide good results inasmuch as the classes are well-separated in the designated feature space.

Blind deconvolution is the basis for a second commonly used approach for classifying $z(t)$: a clean signal $\hat{x}(t)$ is estimated from $z(t)$, then a classifier is used on features of $\hat{x}(t)$. There are many examples of trying to remove multipath by blind deconvolution in order to classify.[9–15] Some researchers exploit the sparseness of the unknown $h(t)$ for producing an estimate $\hat{x}(t)$.[13–15] Broadhead and Pflug report[13] excellent correlation between true signals and signals blindly deconvolved by the minimum entropy method with Cabrelli's sparsity criterion,[16] but did not consider classification. We have shown that these blind deconvolution estimates can be highly correlated to out-of-class training signals, so that nearest neighbor classification on correlation scores performs poorly, particularly at low signal-to-noise ratios.[17,18]

A third approach is to predict the z-space representation of the training signals $\{x_i(t)\}$ using a forward model for the multipath $\hat{h}(t)$. This has the advantage of avoiding deconvolution. A classifier is built using estimated training signals $z_i(t)=x_i(t)*\hat{h}(t)$ for $i=1,\ldots,n$ to classify $z(t)$. The forward model $\hat{h}(t)$ has been based on geometry or physical assumptions.[9,19] Liu *et al.* first proposed an in-channel classifier based on free-field training data.[9] They build a classifier by assuming a finite number of multipath reflections for near-bottom target classification. Dasgupta and Carin classify after accounting for multipath via time-reversal imaging, which requires the geometry and sound speed profile of the channel.[19]

We first proposed that to classify a signal corrupted by unknown multipath, jointly considering deconvolution and classification can lead to better performance than traditional approaches that deconvolve then classify in independent steps.[17] Our method leveraged training data to produce a multipath channel candidate $\hat{h}_i(t)$ for each training signal $x_i(t)$ given $z(t)$. Then, a nearest-neighbor classifier chose the class $y^*=y_i$ for which the estimated filter $\hat{h}_i(t)$ was most multipath-like, according to Cabrelli's sparsity criterion.[16] The resulting joint deconvolution and classification method yields the best signal estimate $\hat{x}(t)=x_i(t)$ and filter estimate $\hat{h}(t)=\hat{h}_i(t)$ that may have produced $z(t)$, as well as the optimal class label $y^*=y_i$. Classification performance was markedly better than minimum entropy blind deconvolution followed by classification, particularly at low signal-to-noise ratios. However, the performance of that joint deconvolution/classifier relied on several conditions.[17] First, it required a good criterion for evaluating how well a given $\hat{h}(t)$ represented a multipath filter. Although Cabrelli's sparsity criterion is an intuitive and convenient choice, real multipath filters can violate the maximal sparsity assumption.[9] Second, the proposed nearest-neighbor approach required that the training signals $\{x_i(t)\}$ be plentiful and that the true $x(t)$ be close to a training sample of the correct class in terms of $\|x(t)-x_i(t)\|$. Third, the deconvolution estimate $\hat{x}(t)$ was always restricted to be a member of the set $\{x_i(t)\}$. Last, it is not straightforward to incorporate features in classification.

II. JOINT MAP DECONVOLUTION AND CLASSIFICATION

A joint deconvolution and classification estimate of the signal, filter, and class label can be constructed using the MAP criterion. Let vectors z, x, h, and w be critically sampled versions of signals $z(t)$, $x(t)$, $h(t)$, and $w(t)$. In this section, we assume that x and w are realizations of random vectors drawn from independent Gaussian distributions. Real signals x certainly may possess non-Gaussian characteristics but the Gaussian assumption is critical to keeping an otherwise formidable deconvolution problem tractable. We assume that the distribution of w is zero-mean with diagonal covariance matrix $\sigma_w^2 I$, where I is the identity matrix; that the probability of x conditioned on the class label y has mean $\mu_{x|y}$ and covariance $\Sigma_{x|y}$; and that the distributions of x, h, and w are mutually independent. We model h using a multivariate Laplacian distribution with independent dimensions, so that the ith element of the random multipath has mean $\mu_h[i]$ and scale parameter $b[i]$. The Laplacian distribution is an appropriate prior model for multipath since it yields sparse realizations. Let $\theta=\{\mu_{x|y},\Sigma_{x|y},\mu_h,b,\sigma_w\}$ be the set of parameters for these three distributions, where θ is assumed to have been modeled or estimated *a priori*.

Then, the proposed joint MAP class estimate y^* solves

$$y^* = \arg\max_y \left(\max_{x,h} p(x,h,y|z,\theta)\right) \tag{2}$$

$$= \arg\max_{y} [\max_{x,h} (p(z|x,h,y,\theta)p(h|\theta)p(x|y,\theta))p(y)] \quad (3)$$

$$= \arg\min_{y} \left[\min_{x,h} \left(\|z - h * x\|^2 + \sigma_w^2 \|x - \mu_{x|y}\|^2_{\Sigma^{-1}_{x|y}} + 2\sigma_w^2 \sum_i \frac{|h_i - \mu_i|}{b_i} \right) + \sigma_w^2 \log|\Sigma_{x|y}| - 2\sigma_w^2 \log p(y) \right], \quad (4)$$

where Eq. (3) follows from Eq. (2) using Bayes' rule, the chain rule, and independence assumptions; and Eq. (4) follows from Eq. (3) by taking the negative logarithm of the pdfs, removing constants that do not depend on x, h, or y from the arg min, and scaling each term by $2\sigma_w^2$. Throughout the paper, we use the notation $\|x\|$ to denote the ℓ_2 norm and $\|x\|_A^2$ for $x^T A x$. The $\|z-h*x\|^2$ term in Eq. (4) drives the estimated filter h and test signal x to be consistent with the received signal z in terms of squared error. The next two terms in Eq. (4) drive x to match the *a priori* expected signal via the ℓ_2 norm and drive h to match the *a priori* expected filter via the ℓ_1 norm. Note that these latter two terms are regularized by the noise variance—the greater the noise power, the more the estimate relies on the *a priori* expectations and less on matching the received signal $z(t)$. The fourth term penalizes classes that exhibit high variance, and the fifth term is the class membership prior. Since the noise determines the degree of regularization, a curious behavior of this approach is that it performs poorly for high SNR: the first term will dominate as σ_w^2 goes to zero, and solutions for x and h will no longer depend on $\mu_{x|y}$ and μ_h, respectively.

The objective function in Eq. (4) is not convex since it involves a product of variables in the convolution integral $h*x$. However, the problem is jointly convex in h and x in the limit as $\sigma_w \to \infty$, and is marginally convex in x or h for all σ_w. Therefore, we opt to solve Eq. (4) using an alternating minimization approach as a heuristic for finding the true solution. Using H for the Toeplitz matrix representation of discrete convolution with fixed h,[20] and for fixed y, the objective as a function of x can be written in the form of generalized Tikhonov regularization $\|Hx-z\|^2 + \sigma_w^2 \|h-\mu_h\|^2_{\Sigma_h^{-1}}$. The solution[21] is

$$\hat{x} = (H^T H + \sigma_n^2 \Sigma_{x|y}^{-1})^{-1} (H^T z + \sigma_w^2 \Sigma_{x|y}^{-1} \mu_{x|y}).$$

Next, we solve Eq. (4) for those terms depending on h by fixing x and rewriting as

$$\hat{h} = \arg\min_h \|Xh - z\|_2^2 + \sigma_w^2 \|D^{-1}(h - \mu_h)\|_1, \quad (5)$$

where X is the Toeplitz matrix representation of discrete convolution of x with h, D is a diagonal matrix with entries $b_i/2$, and $\|\cdot\|_1$ is the ℓ_1 norm. Equation (5) can reformulated as a quadratic program with linear constraints.[22]

Since the optimization problem in Eq. (4) is non-convex, the alternating minimizations strategy is not guaranteed to converge to the global minimum.[23] A common approach is to optimize starting from several initial points, then choose the overall minimizer. The initial guesses could be drawn i.i.d. from the class-conditional distribution $\mathcal{N}(\mu_{x|y}, \Sigma_{x|y})$. We use a slightly different approach to take advantage of the fact that we have examples from each class: convex combinations of the training signals are initial guesses.

Experiments and results for the joint MAP classifier are presented in Sec. III.

A. A related MAP deconvolution approach

MAP deconvolution has been explored previously by Lam and Goodman for blind image deblurring (without classification).[24] In that work, Lam and Goodman estimate the point spread function h and the image covariance Σ_x by maximizing $p(z|h,\Sigma_x)p(h)p(\Sigma_x)$. The prior $p(\Sigma_x)$ is replaced with a heuristic smoothness constraint on the covariance, and the prior $p(h)$ is replaced with the hard constraint $h \in \mathcal{H}$ for some convex set $\mathcal{H}$. They propose an EM algorithm implementation that alternates between estimating Σ_x (the E-step) and h (the M-step) in the Fourier domain. The image is finally estimated by Wiener deconvolution using the estimated h and Σ_x. The algorithm results in high-quality deblurred image estimates.[24]

The Lam and Goodman MAP blind deconvolution may be extended to the multipath problem by applying a Laplacian prior for $p(h)$ as we have done in Eq. (4) instead of their hard constraint $h \in \mathcal{H}$. However, their approach cannot be extended to fit in the joint deconvolution/classification paradigm by simply conditioning on the class label y and adding the prior $p(y)$ in the optimization. First, Lam and Goodman assume that the image x (and therefore, z) is a realization of a zero-mean Gaussian distribution. In our framework, the class-conditional mean is an important discriminating feature of the class. Second, their method must estimate Σ_x, but in our framework the class conditional covariance $\Sigma_{x|y}$ is estimated *a priori* from training pairs. Naively replacing Σ_x with $\Sigma_{x|y}$ renders their E-step useless so that iterating does not improve the initial guess. Thus, the training pairs $\{x_i, y_i\}$ offer little advantage to their MAP blind deconvolution technique.

III. PROBABILISTIC DECONVOLUTION AND CLASSIFICATION USING QDA

Estimating the true signal is difficult and unnecessary if only a class label is required. In this section, we explore classifying signals jointly with probabilistic deconvolution, in which a statistical characterization of $x(t)$ and $h(t)$ are used without ever choosing a particular, deterministic signal or channel estimate. Specifically, we consider the maximum likelihood classifier that solves

$$y^* = \arg\max_y p(z|y) = \arg\max_y \int\int p(z|x,h,y)p(x|y)p(h)p(y)\, dx\, dh. \quad (6)$$

Assuming a uniform prior, the classifier in Eq. (6) differs from the joint MAP classifier in Eq. (2) in that the $\max_{x,h}$ operator in Eq. (2) is replaced by expectation over x and h in Eq. (6). For the remainder of this paper, we will assume uniform prior $p(y)$ such that $\arg\max_y p(y|z) = \arg\max_y p(z|y)$.

QDA is a popular classification rule that models each class-conditional distribution in Eq. (6) as Gaussian.[25–27] This assumption can be motivated by the central limit theorem and the fact that the Gaussian is the maximum entropy (least assumptive) distribution given first and second moments. Here, we build a QDA classifier in z-space by assuming $p(z|y)$ in Eq. (6) is Gaussian, and we show that one can calculate the sufficient statistics $\mu_{z|y}$ and $\Sigma_{z|y}$ of $p(z|y)$ from the estimated mean and covariance of the channel $h(t)$ and the estimated means and covariances of the training signals from each class. Note that we do not make any assumptions on the distributions of h or x given y other than that they have finite first and second moments; in fact, the result of the convolution $h*x$ would not be Gaussian if h and x were assumed to be realizations of Gaussian processes. This section expands on a recent workshop paper.[18]

A. QDA classification of signals corrupted by LTI filtering

Let x be a realization of a random process with finite class-conditional mean $\mu_{x|y}$ and finite covariance $\Sigma_{x|y}$; let the noise w be a realization of a zero-mean random process with covariance $\sigma_w^2 I$; and let h be a realization of a random process with mean μ_h and covariance Σ_h. Given these statistics, the full derivations of the first and second moments of the distribution of z are given in the Appendix. In summary, the first moment $\mu_{z|y}$ can be written in terms of the mean class-conditional signal and mean filter as

$$\mu_{z|y} = \mu_{x|y} * \mu_h. \tag{7}$$

Likewise, the covariance $\Sigma_{z|y}$ conditioned on class y can be expressed in terms of the class-conditional signal and filter statistics as

$$\Sigma_{z|y} = (\Sigma_{x|y} + \mu_{x|y}\mu_{x|y}^T) ** (\Sigma_h + \mu_h\mu_h^T) + \sigma_w^2 I - \mu_{z|y}\mu_{z|y}^T, \tag{8}$$

where $**$ denotes two-dimensional discrete convolution.

B. Experiments with signal-based joint QDA and joint MAP classification

We tested the proposed methods with two experiments that differ in how we generate the simulated multipath. In the first we use a Laplacian random process to generate realizations of multipath channels. In the second, a random K-sparse model is used. In both experiments the clean, x-space signals are drawn from class-conditional Gaussian distributions, which means that the received signals in z-space are not actually Gaussian distributed.

1. Signal classification experiment: Laplacian multipath

Each coefficient of a multipath filter was drawn independently from a Laplacian random process with parameters $\mu[n]$, $b[n]$,

TABLE I. Simulation parameters for joint MAP/joint QDA experiments. Note that square(n)=sgn(sin(n)).

Parameter	Class 1	Class 2
	Well-separated means	
$\mu_{x\|y}[n]$	$\frac{1}{4}$ square$(6\pi n/100)$	$\frac{1}{4}$ square$(12\pi n/100)$
$\Sigma_{x\|y}[m,n]$	$\frac{1}{100}(\delta[m-n]+e^{-\|m-n\|/20})$	$\frac{1}{100}(\delta[m-n]+e^{-(m-n)^2/10})$
	Close means	
$\mu_{x\|y}[n]$	$\frac{1}{4}$ square$(6\pi n/100)$	$\frac{1}{4}\sin(6\pi n/100)$
$\Sigma_{x\|y}[m,n]$	$\frac{1}{100}(\delta[m-n]+e^{-\|m-n\|/20})$	$\frac{1}{100}(\delta[m-n]+e^{-(m-n)^2/10})$

$$p(h[n]|\mu[n],b[n]) = \frac{1}{2b[n]}e^{-(|x-\mu[n]|)/b[n]},$$

for $n=0,\ldots,99$, where we set the expected filter to be $\mu_h[n]=\delta[n]-0.6[n-49]+0.1\delta[n-99]$, and the scale parameter $b[n]$ decays as n grows: $b[n]=0.2e^{-0.024n}$. The decay parameter coefficients for this experiment were chosen to model oceanic multipath filtering of sonar signals.

Test and training signals were drawn i.i.d. from a Gaussian distribution $\mathcal{N}(\mu_{x|y},\Sigma_{x|y})$ where the class y was drawn uniformly between two classes. We considered two classification scenarios to test performance: classes whose mean vectors are well-separated, and close. The mean signals are composed of square and sine waves, and the covariance matrices are Toeplitz with smooth covariance structure. The specific values of $\mu_{x|y}$ and $\Sigma_{x|y}$ for each experiment are shown in Table I. Each test signal z was created by convolving a randomly drawn signal x with randomly drawn multipath h, and adding Gaussian white noise w to achieve different SNRs, where the SNR is with respect to the multipath signal, $20\log\|x*h\|/\sigma_w$.

We compared the joint QDA classifier to a matched filter that ignores multipath. For the matched filter, the received signal z is tested against $\mu_{x|y}$ for each class. The joint MAP classifier in Eq. (4) is compared to a matched filter on a blind deconvolution signal estimate. For deconvolution, the received signal z is first denoised by Wiener filtering, then $\hat{h}$ is estimated using Cabrelli's blind deconvolution method for signals that have undergone unknown multipath filtering.[16] The estimate $\hat{x}$ is then computed via deconvolution in the Fourier domain. The true signal length was used as a required input to Cabrelli's method. Each of the methods used the true signal and channel statistics, and the true SNR where needed.

2. Signal classification experiment: K-sparse multipath

The K-sparse experiments are the same as described in the previous subsection, except the multipath filters were generated using a sparse model:

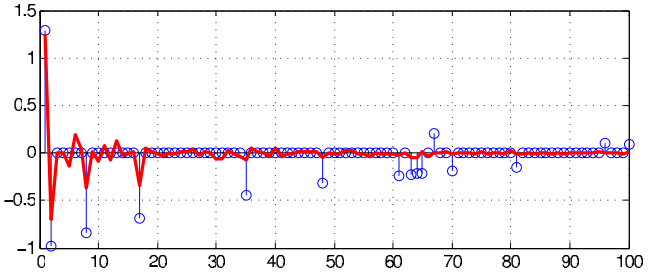


FIG. 1. (Color online) Example multipath realization from the K-sparse model (stem), and the deconvolution estimate produced by the joint MAP deconvolution/classifier (solid) at 10 dB SNR.

$$h[n] = \sum_{i=1}^{K} \alpha_i \delta[n - d_i],$$

with $K=15$ nonzero coefficients, delays d_i drawn uniformly on $[0,99]$, $\alpha_i = \pm e^{-\beta d_i}$ with randomly chosen sign, and decay parameter $\beta=0.0240$ chosen to mimic real underwater acoustic channels. An example realization of a filter h drawn from this model is shown in Fig. 1. The diagonal covariance matrix Σ_h is estimated from 1000 samples of the impulse response.

C. Signal-based joint QDA and joint MAP results

Figure 1 shows a reconstructed multipath estimate produced by the joint MAP deconvolution/classifier corresponding to the chosen class for the well-separated means experiment at 10 dB SNR. In this case, joint MAP correctly identified the class label. The recovered filter is a reasonable reconstruction of the true filter, but generally underestimates the amplitude of the first coefficients, and does not reliably reconstruct the tail of h. The gross errors can be ascribed to the fact that the optimization problem in Eq. (4) is not convex, and to the mismatch between the Laplacian prior and K-sparse model.

Classification results in Fig. 2 show that the proposed joint QDA classifier dominates the matched filter classifier for both Laplacian multipath in (a) and (b), and for multipath generated by the K-sparse model in (c) and (d). The means for each class used for (a) and (c) (well-separated means) are orthogonal, so the matched filter performs well despite ignoring the multipath. However, for (b) and (d) the means are similar, and the matched filter performs poorly compared to joint QDA. The joint MAP classifier performs well at low SNR, but as predicted, performance degrades as SNR increases. For truly sparse multipath in (c) and (d), the joint MAP approach is unaffected for well-separated means in (c), and affected moderately at high SNR for close means in (d) compared to results using Laplacian multipath. Evidently, the ℓ_1 is an appropriate heuristic for the K-sparse multipath model; using ℓ_1 criteria to obtain sparse solutions is a popular approach.[28]

IV. QDA CLASSIFICATION OF SIGNALS CORRUPTED BY LTI FILTERING USING FEATURES

Blind signal deconvolution and the proposed joint MAP and signal-based joint QDA methods are computationally

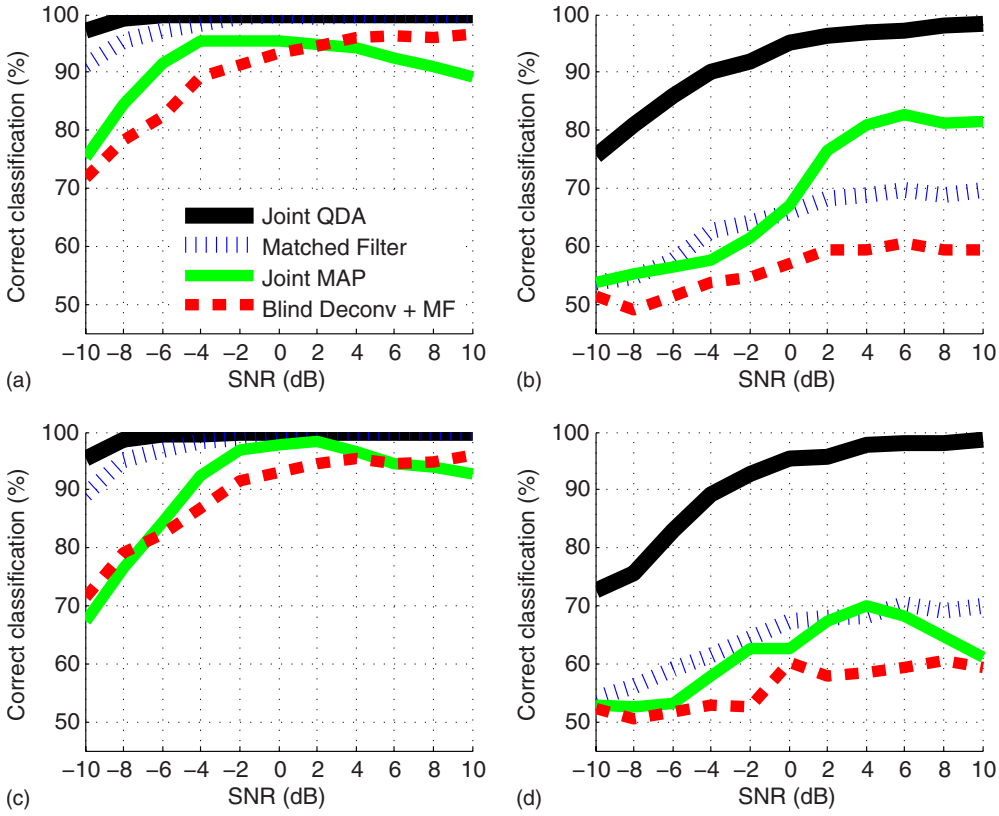


FIG. 2. (Color online) Classification accuracy for four experiments using multipath generated from a Laplacian model in (a) and (b), and a K-sparse model in (c) and (d). The results are averaged over 1000 i.i.d. test signals for each SNR point.

prohibitive for signals captured at high sample rates. For M-length sampled signals, Cabrelli's method requires the inversion of an $M \times M$ Toeplitz matrix, which at best is of complexity $O(M \log M)$. The joint MAP deconvolution requires more iterations to converge as M increases, and requires (as does joint QDA) inversion of an $M \times M$ covariance matrix, which in general is complexity $O(M^3)$. To decrease the computational burden and possibly increase classification performance, an alternative is to classify based on features that represent important characteristics of the signals and provide good class discrimination.[29] The hope is that classes can be well-discriminated by features of significantly smaller dimensionality $d \ll M$.

Unless multipath-invariant features are used, a classifier trained on x-space features will not generally be applicable to classify the features of $z(t)$ directly. However, if a functional relationship can be found that relates x-space features to their images in z-space, then a suitable classifier can be trained and applied in z-space. We show that such a relationship can be derived for subband power features, which represent an important class of discriminating features for many remote-sensing applications. Extending the proposed joint QDA classifier to subband features in z-space requires expressing the mean and covariance of the subband power feature vector for $z(t)$ in terms of statistics of the subband power features of the channel and training data.

Let the d-dimensional feature vector $P_z = [P_z(f_1) \cdots P_z(f_d)]^T$ be composed of the subband powers of $z(t)$ at frequencies $\{f_i\}$ for $i = 1, \ldots, d$. Then for any frequency f, because $P_z(f) = Z(f)Z^*(f)$, and $Z(f) = X(f)H(f) + W(f)$, the subband power can be expressed as

$$P_z(f) = P_x(f)P_h(f) + P_w(f) + 2\,\mathrm{Re}\{X(f)H(f)W^*(f)\}, \quad (9)$$

where $P_x(f)$, $P_h(f)$, and $P_w(f)$ denote the power of the signal, the channel, and the noise for frequency f, respectively. Based on Eq. (9), derivations for the mean and covariance of the feature vector P_z are given in the Appendix. For these derivations, it is assumed that $w(t)$ is a realization of a Gaussian white noise process. The class-conditional mean feature vector $\mu_{P_z|y}$ can be expressed in terms of the noise power σ_w^2 and mean vectors of the clean signal features and channel features, $\mu_{P_x|y}$ and μ_{P_h}, respectively, as

$$\mu_{P_z|y} = \mu_{P_x|y} \cdot \mu_{P_h} + \sigma_w^2 \mathbf{1}, \quad (10)$$

where $\cdot$ denotes Hadamard (Schur or element-wise) multiplication, and $\mathbf{1}$ is a vector of ones. The class-conditional covariance $\Sigma_{P_z|y}$ can be expressed in terms of σ_w^2 and second-order statistics of $P_x|y$ and P_h as

$$\Sigma_{P_z|y} = \Sigma_{P_x|y} \cdot \Sigma_{P_h} + \sigma_w^4 I + \Sigma_{P_h} \cdot \mu_{P_x|y}\mu_{P_x|y}^T + \Sigma_{P_x} \cdot \mu_{P_h}\mu_{P_h}^T + 2\sigma_w^2\,\mathrm{diag}\{\mu_{P_x|y} \cdot \mu_{P_h}\}. \quad (11)$$

It is assumed that μ_{P_h}, Σ_{P_h}, and σ_w^2 can be estimated from the channel. The statistics $\mu_{P_x|y}$ and $\Sigma_{P_x|y}$ are estimated from training data for each class. Together, these statistics are used to compute class-conditional QDA model parameters $\mu_{P_z|y}$ and $\Sigma_{P_z|y}$ from Eqs. (10) and (11) and, hence, build a QDA subband power feature classifier in the z-space.

A. Informed and blind classifiers using features

To evaluate the proposed z-space joint QDA using power features, we consider two alternate approaches to classification from power features. First, an "informed" x-space classifier approach uses the expectation of the channel's subband power response $E[P_h]$ and the noise power σ_w^2 to transform the z-space features to x-space by subtracting the noise power and deconvolving by $E[P_h]$: $\hat{P}_x(f_i) = (P_z(f_i) - \sigma_w^2)/E[P_h(f_i)]$. This approach uses the mean but not covariance of the channel.

Second, a "blind" x-space classifier approach uses x-space features, and simply ignores statistics of the channel and noise altogether. Training signals are first normalized by their total power before a classifier is trained. Then, a received signal $z(t)$ is also normalized by its total power before extracting features.

We compare joint QDA to both blind and informed approaches for QDA, 1-NN, and a support vector machine (SVM).[25]

B. Simulated classification experiments using joint QDA with power features

We consider the task of classifying narrow-band signals corrupted by unknown multipath due to propagation in a

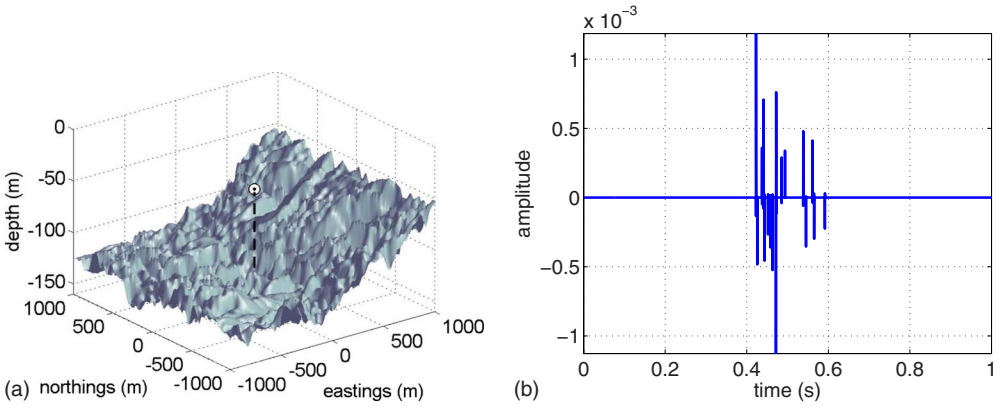


FIG. 3. (Color online) (a) Simulated ocean bathymetry with a single receiver (marked by ⊙) at (0,0,−50) m, and (b) a sample channel impulse response for a source located at (460,250,−70) m, generated by the Sonar Simulation Toolset (Refs. 30 and 31).

TABLE II. Pole magnitude distribution for feature-based classification experiments.

Parameter	Class 1	Class 2
$\Sigma_{a\|y}$	$\begin{pmatrix}1.00 & 0.99\\ 0.99 & 9.00\end{pmatrix}\times 10^{-6}$	$\begin{pmatrix}6.00 & -0.80\\ -0.80 & 1.00\end{pmatrix}\times 10^{-4}$
	Close poles	
$\mu_{a\|y}$	(0.945 0.905)	(0.909 0.948)
	Moderately-separated poles	
$\mu_{a\|y}$	(0.945 0.875)	(0.879 0.948)
	Well-separated poles	
$\mu_{a\|y}$	(0.965 0.875)	(0.875 0.948)

shallow ocean channel. To simulate two classes of narrowband signals with two subband power features, training and test signals are generated i.i.d. using the z-domain model

$$X|y[z] = \frac{(z-1)(z+1)}{\Pi_{\ell=1}^{2}(z-p_{\ell|y})(z-p_{\ell|y}^{*})}, \tag{12}$$

where the location of each class-conditional pole $p_{\ell|y}$, $\ell \in \{1,2\}$ is drawn randomly from the model $a_{\ell|y}\exp(j\theta_\ell)$, where θ_ℓ is fixed, and $a_{\ell|y}=[a_{1|y},a_{2|y}]^T$ is multivariate Gaussian distributed with mean $\mu_{a|y}$ and covariance matrix $\Sigma_{a|y}$. Although the vector a for each class is Gaussian distributed, the signals in feature space are not, as per Eq. (12). Figure 4 shows an example pole-zero plot and corresponding log-feature space scatterplot for a well-separated case. We consider three instances of the experiment for choices of $\mu_{a|y}$ that result in different class separation. The parameters $\mu_{a|y}$ and $\Sigma_{a|y}$ for each instance of the experiment are shown in Table II, and $\theta_1=\frac{1}{50}$ and $\theta_2=\frac{1}{5}$. Note that since all poles and zeros lie within the unit circle, for each case the selected parameters correspond to a realization of a minimum phase signal, which could be produced from natural sources.

Test and training signals were generated by taking i.i.d. draws of poles as described above, and taking 5000 evenly spaced samples around the unit circle of the complex z-plane, so that the length of each signal corresponds to 1.25 s, sampled at 4 kHz. At frequencies θ_1 and θ_2 the subband power is extracted from each signal and used as classification features. The parameters in Table II were chosen such that the generated test and training signals were linearly separable in the subband power feature space.

Channel impulse responses h were drawn i.i.d. in the following manner. A receiver is placed at a depth of 50 m in a simulated shallow water channel, as shown in Fig. 3. Source locations were drawn uniformly from the cube 2 km across north and east and 150 m deep; locations falling below the ocean floor are discarded and redrawn. Channel impulse responses h were generated by propagating an impulsive source from the random source locations to the receiver using the *CASS Eigenray* routine provided in the Sonar Simulation Toolset.[30] Impulse responses were sampled at 4 kHz. The ocean environment is set up to be fairly extreme, but static. We have imposed a prototypical sound speed profile (ranging from 1477 to 1492 m/s), and have modeled the ocean bottom to contain sandy gravel with mean grain size 2 mm. Surface roughness is governed by the wind speed, which is set to 15 km/h. The channel geometry and a sample channel impulse response are shown in Fig. 3.

Maximum likelihood estimates of $\mu_{P_x|y}$ and $\Sigma_{P_x|y}$ were computed from 1000 randomly drawn training signals, and maximum likelihood estimates of μ_{P_h} and Σ_{P_h} were computed from 1000 randomly drawn channel impulse responses. The test samples were corrupted with randomly drawn multipath, and then i.i.d. white noise with variance σ_w^2 was added, where the multipath-corrupted-signal to noise ratio was varied between −10 and 10 dB. Classification results were averaged over 5000 trials for each SNR.

We compare the proposed z-space joint QDA classifier to x-space QDA, SVM with a linear kernel, and 1-NN classifiers using both the informed and blind approaches described in Sec. IV A. In each of the three simulations, the training and test data were linearly separable. Therefore, we set the regularization term C in the linear SVM (Ref. 32) to a large value, $C=10^7$. Then, none of the classifiers requires cross-validation.

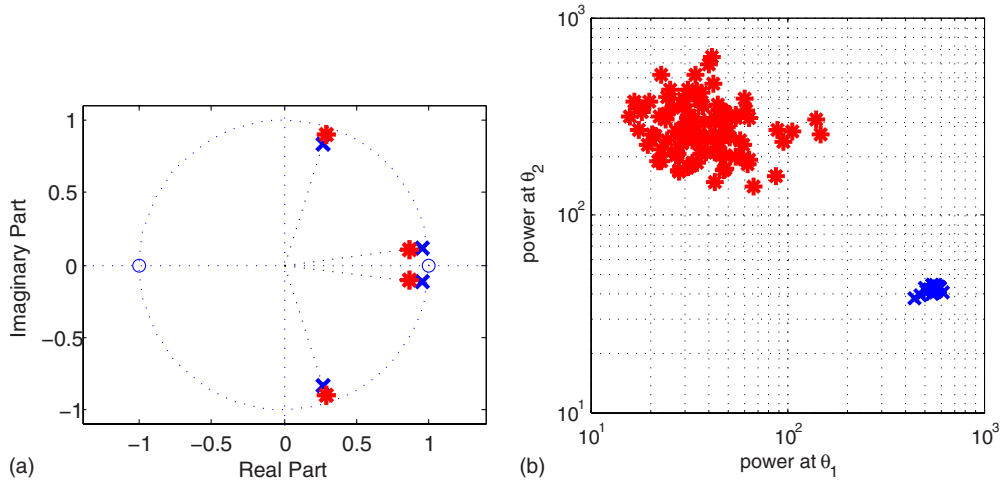


FIG. 4. (Color online) (a) Pole-zero plot showing the mean location of the poles for each class 1 (×) and class 2 (∗) for the "well-separated" case, and (b) scatter-plot of the classes in log-feature space.

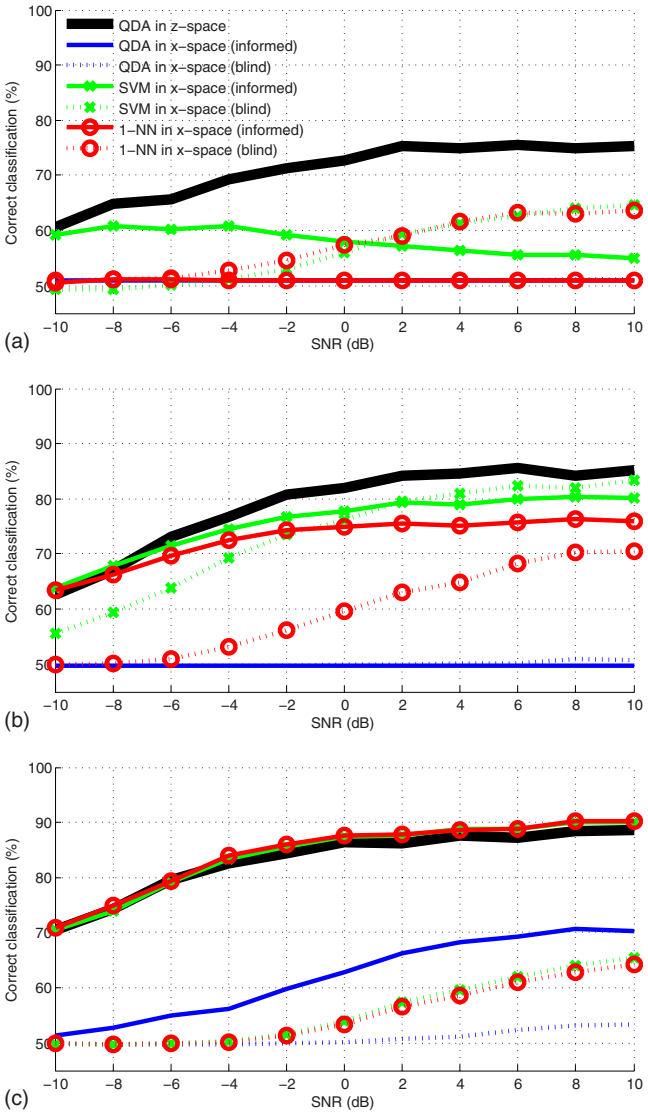


FIG. 5. (Color online) Results for feature-based classification on simulated data for the experiments where classes are (top) close in feature space, (middle) moderately separated in feature space, and (bottom) well-separated in feature space.

C. Simulation results

Results for each experiment are shown in Fig. 5. The joint QDA method using features (QDA in z-space) performs markedly better than the other approaches when the classes are difficult to separate. As class separation increases, the performance of the z-space QDA and informed SVM and 1-NN become similar.

D. Classifying bowhead whale songs in multipath channel

We employ the feature-based joint QDA method to identify individual Bowhead whales in a multipath environment by classifying the end notes of their songs. Several end notes of Bowhead whale vocalizations for two individuals were extracted from the MobySound archive.[33] Fifteen vocalizations are available for whale 1, and nine vocalizations are available for whale 2. According to the metadata, the end notes of Bowhead whale songs are relatively stable from year to year. Therefore, we hope to be able to acoustically discriminate between two individuals based on previously recorded vocalizations. Our experimental setup simulates a shallow ocean channel (in comparison to the observation distances), and low SNR. Each of the signals has non-negligible interfering noise from bearded seals, sea ice and banging hydrophone cables.[33] The vocalizations were recorded in April 1988 near the coast of Point Barrow, Alaska, but for these experiments, we inject the signals into randomly drawn locations in the simulated bathymetry shown in Fig. 3(a). Example vocalizations for each whale are shown in Fig. 6.

Classifiers were trained on five training signals drawn at random for each class. The remaining 14 signals were propagated from a random source location in the bathymetry to the receiver using the *CASS Eigenray* routine. Gaussian white noise is added to the multipath signal to achieve a specified SNR. To increase the statistical significance of the results, classification results were averaged over 1000 iterations of the random training/test partitioning with random source locations.

For classification, four peak power features were selected: the two largest amplitude peaks averaged over signals in class 1 (163 and 258 Hz) and the two largest amplitude peaks for class 2 (588 and 207 Hz). Features do not correspond to strong interfering noise, and result in classes that, prior to the channel effects, are 100% linearly separable. As before, we compare the z-space joint QDA classifier to x-space QDA, linear SVM, and 1-NN classifiers using both informed and blind approaches.

E. Bowhead whale song results

Results for the Bowhead whale songs are shown in Fig. 7. The z-space joint QDA classifier using power features consistently achieves roughly 4% higher accuracy than other methods across the range of SNRs. This can be ascribed to the fact that the multipath channel distorts the relative covariance structures of class 1 and class 2 whale vocalizations.

V. CONCLUSIONS AND OPEN QUESTIONS

We have presented classification methods that jointly consider the effects of multipath distortion with classification. In particular, we have investigated a joint MAP deconvolution/classifier that incorporates first- and second-order statistics of the channel and yields a MAP solution for the recovered signal $\hat{x}$, the recovered filter $\hat{h}$, and the class estimate y^*. Two drawbacks of the joint MAP algorithm are that it is not convex, and that it theoretically performs poorly at high SNR. The first problem might be addressed by maximizing the marginal $p(h,y|z)$ which yields a convex expression, but requires more complicated optimization approaches. The second problem arises since regularization scales with the noise power σ_w^2, which may be replaced by a fixed penalty that can be chosen via cross-validation.

We hypothesized that better classification performance can be gained by marginalizing over x and h. To that end, we presented a joint QDA classifier that accounts for the LTI corruption probabilistically. Experiments showed that the joint QDA classifier outperformed the joint MAP classifier

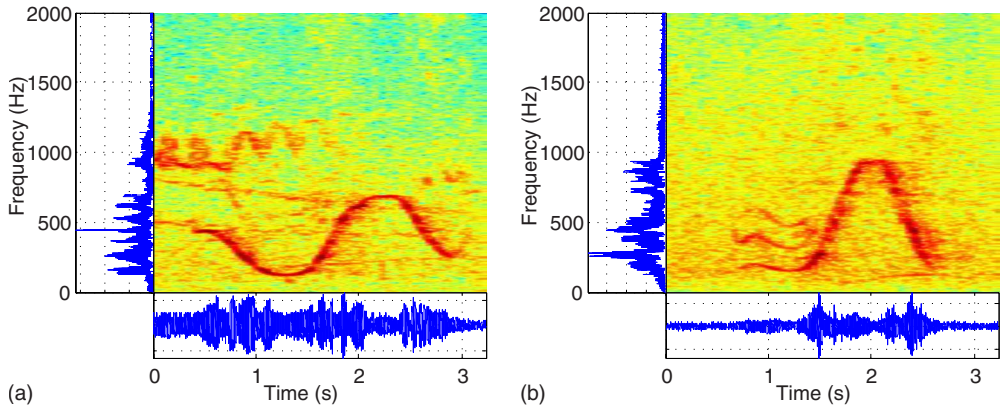


FIG. 6. (Color online) Spectrograms of whale song-endnotes for (a) the first Bowhead whale and (b) the second Bowhead whale. The vocalizations of the second whale tend to be more variable, cover a greater dynamic range, and contain stronger harmonic components than the first whale. Notice that the vocalization in (a) contains interfering calls from a bearded seal from about 800 to 1200 Hz.

and a classifier based on blind deconvolution. Further, we derived a joint QDA classifier that classifies based on subband power features. Experiments on both simulated signals and real Bowhead whale vocalizations demonstrated the efficacy of the joint QDA classifier using features. Although we have derived the joint QDA classifier for subband power features, there are many other important classes of features for signal processing including real cepstrum, wavelets, and wavepacket decomposition.[12] In cases where the features are linear functions of the data, such as the shift invariant wavepackets described in Ref. 12, the derivation of a joint QDA classifier is straightforward. However, features that are nonlinear functions of the received signal (e.g., real cepstrum) may require low-order approximate solutions.

We compared the joint QDA classifier on subband features to other classifiers that also took into account the mean multipath using subband power features. We showed that using this first-order multipath information significantly improved performance over ignoring the multipath altogether, except in the simulation experiment with well-separated classes and low noise. An open question in this line of research is how to take advantage of further information about the multipath for other classifiers, such as the nearest-neighbor and support vector machine classifiers.

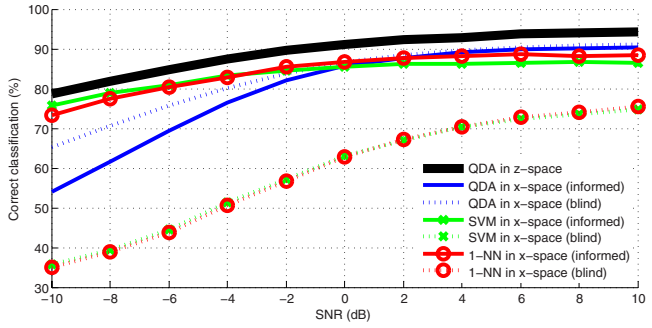


FIG. 7. (Color online) Classification results for identifying Bowhead whales by the end-notes of their songs. The SVM in x-space (blind) and 1-NN in x-space (blind) mostly overlap one another. Note that unlike previous figures, the ordinate axis shows results down to 30% correct classification to better show the trend over SNR.

APPENDIX: DERIVATIONS

In the Appendix, we use x, h, w, and z (and hence, all functions of them, for example P_x) to denote random signals, not their realizations as done in the main body.

Derivation of the first and second moments of $z(t)$

This section details the derivation of Eqs. (7) and (8). Since the noise is zero-mean, the nth component of the mean time signal $\mu_{z|y}$ is

$$\mu_{z|y}[n] = E[(x*h)[n] + w[n]|y] = E\left[\sum_k x[k]h[n-k]|y\right]$$

$$= \sum_k E[x[k]|y]E[h[n-k]] = \sum_k \mu_{x|y}[k]\mu_h[n-k],$$

and thus $\mu_{z|y} = \mu_{x|y} * \mu_h$.

The covariance matrix is derived:

$$\Sigma_{z|y} = E[zz^T|y] - E[z|y]E[z|y]^T$$

$$\overset{(a)}{=} E[(x*h)(x*h)^T|y] + E[ww^T] - \mu_{z|y}\mu_{z|y}^T$$

$$\overset{(b)}{=} E[(xx^T) ** (hh^T)|y] + \sigma_w^2 I - \mu_{z|y}\mu_{z|y}^T$$

$$\overset{(c)}{=} E[xx^T|y] ** E[hh^T] + \sigma_w^2 I - \mu_{z|y}\mu_{z|y}^T,$$

where the expectations are taken with respect to the appropriate distributions. In the above, (b) follows from (a) because the (m,n)th component of the outer product $(x*h) \times (x*h)^T$ can be expressed as

$$\left(\sum_k x[k]h[n-k]\right)\left(\sum_i x[i]h[m-i]\right)$$

$$= \sum_k \sum_i x[k]h[n-k]x[i]h[m-i]$$

$$= \sum_k \sum_i (x[k]x[i])(h[n-k]h[m-i]),$$

and thus $(x*h)(x*h)^T = (xx^T) ** (hh^T)$. Then, (c) follows from (b) due to the linearity of both expectation and convolution. Finally, (c) can be rewritten as Eq. (8).

Derivation of the first and second moments of $P_z(f)$

This section details the derivation of Eqs. (10) and (11). Let $\mu_{P_z|y}=E[P_z|y]$, $\mu_{P_x|y}=E[P_x|y]$, $\mu_{P_h}=E[P_h]$ and $E[P_w]=E[WW^*]=\sigma_w^2\mathbf{1}$ by assumption, where $\mathbf{1}$ is a vector of ones. The first moment given by Eq. (10) follows from Eq. (9) since $E[\mathrm{Re}\{a\}]=E[\frac{1}{2}(a+a^*)]=\frac{1}{2}(E[a]+E[a]^*)=\mathrm{Re}\{E[a]\}$, and $X(f)$, $H(f)$, and $W^*(f)$ are independent, so $E[\mathrm{Re}\{X(f)H(f)W^*(f)\}]=0$ because $E[W^*(f)]=0$. The fact that $E[W^*(f)]=0$ follows from $E[w(t)]=0$ and the independence of the noise random process over time.

For the class-conditional covariance $\Sigma_{P_z|y}$, we derive the second moment and cross-correlation separately. For notational simplicity, we denote $E[x|y]$ by $E[x]$, that is, the class-conditional membership is implied in the expectation. For the second moment, it follows from Eq. (9) that

$$
\begin{aligned}
E[P_z(f)^2] &\overset{(a)}{=} E[P_x^2P_h^2+P_w^2+4\,\mathrm{Re}(XHW^*)^2 \\
&\quad +4(P_xP_h+P_w)\mathrm{Re}(XHW^*)+2P_xP_hP_w] \\
&\overset{(b)}{=} E[P_x^2]E[P_h^2]+E[P_w^2]+4E[\mathrm{Re}(XHW^*)^2] \\
&\quad +2E[P_x]E[P_h]E[P_w] \\
&\overset{(c)}{=} E[P_x^2]E[P_h^2]+E[P_w^2]+2E[XX^*HH^*WW^* \\
&\quad +\mathrm{Re}((XHW^*)^2)]+2E[P_x]E[P_h]E[P_w] \\
&\overset{(d)}{=} E[P_x^2]E[P_h^2]+E[P_w^2]+4E[P_x]E[P_h]E[P_w] \\
&\quad +2\,\mathrm{Re}(E[XX]E[HH]E[W^*W^*]) \\
&\overset{(e)}{=} E[P_x^2]E[P_h^2]+E[P_w^2]+4E[P_x]E[P_h]E[P_w]. \qquad \text{(A1)}
\end{aligned}
$$

In the above, (b) follows from (a) because $4E[(P_xP_h+P_w)\mathrm{Re}(XHW^*)]$ can be expanded into the sum of two terms containing $E[W^*]$ and $E[W^*WW^*]$, and since these terms equal zero, $4E[(P_xP_h+P_w)\mathrm{Re}(XHW^*)]=0$. For (c), we employed the identity $(\mathrm{Re}\{c\})^2=1/2(cc^*+\mathrm{Re}(c^2))$. Then (d) follows from (c) by the definition of power and the interchangeability of expectation and the real operator, explained earlier in this section. The step from (d) to (e) holds because the Fourier transform $W(f)=\mathcal{F}[w(t)]$ of a zero-mean Gaussian white noise process is a complex zero-mean Gaussian white process with $\mathrm{Re}(W)$ and $\mathrm{Im}(W)$ uncorrelated, $E[\mathrm{Re}(W)]=E[\mathrm{Im}(W)]=0$, and $E[\mathrm{Re}(W)^2]=E[\mathrm{Im}(W)^2]=\frac{1}{2}\sigma_w^2$, so that $E[W^*W^*]=E[\mathrm{Re}(W)^2-2j\,\mathrm{Re}(W)\mathrm{Im}(W)-\mathrm{Im}(W)^2]=0$.[34]

Last, we derive the cross-correlation using as short-hand P_{z_i} to denote $P_z(f_i)$, P_{h_i} to denote $P_h(f_i)$, H_i to denote $H(f_i)$, etc. From Eq. (9), for $i\neq j$,

$$
\begin{aligned}
E[P_{z_i}P_{z_j}] &= E[P_{x_i}P_{h_i}P_{x_j}P_{h_j}]+E[P_{x_i}P_{h_i}P_{w_j}] \\
&\quad +E[P_{x_j}P_{h_j}P_{w_i}]+E[P_{w_i}P_{w_j}]+2E[\mathrm{Re}(X_jH_jW_j) \\
&\quad \times(P_{x_i}P_{h_i}+P_{w_i})]+2E[\mathrm{Re}(X_iH_iW_i)(P_{x_j}P_{h_j} \\
&\quad +P_{w_j})]+4E[\mathrm{Re}(X_iH_iW_i)\mathrm{Re}(X_jH_jW_j)], \qquad \text{(A2)}
\end{aligned}
$$

Consider the fifth term of Eq. (A2),

$$
\begin{aligned}
2E[\mathrm{Re}(X_jH_jW_j)(P_{x_i}P_{h_i}+P_{w_i})] &= E[(X_jH_jW_j+X_j^*H_j^*W_j^*) \\
&\quad \times(X_iX_i^*H_iH_i^*+W_iW_i^*)].
\end{aligned}
$$

After multiplication, the above consists of four terms, each containing a single W_j. Since W_j is uncorrelated with every other random variable and $E[W_j]=0$, each of the four terms in the expansion has zero mean, and thus the entire expression is zero. By the same logic, the sixth term of of Eq. (A2) is

$$
2E[\mathrm{Re}(X_iH_iW_i)(P_{x_j}P_{h_j}+P_{w_j})]=0.
$$

The last term of Eq. (A2) can be rewritten

$$
\begin{aligned}
4E[\mathrm{Re}(X_iH_iW_i)\mathrm{Re}(X_jH_jW_j)] &= E[(X_iH_iW_i+X_i^*H_i^*W_i^*) \\
&\quad \times(X_jH_jW_j+X_j^*H_j^*W_j^*)].
\end{aligned}
$$

Taking the product results in four terms which each have a single W_i and W_j. Since W_i and W_j are uncorrelated and each has zero-mean, each of the four terms is zero. Thus, the cross-correlation of any two frequencies $f_i\neq f_j$ is

$$
\begin{aligned}
E[P_{z_i}P_{z_j}] &= E[P_{x_i}P_{x_j}]E[P_{h_i}P_{h_j}]+E[P_{x_i}]E[P_{h_i}]E[P_{w_j}] \\
&\quad +E[P_{x_j}]E[P_{h_j}]E[P_{w_i}]+E[P_{w_i}]E[P_{w_j}]. \qquad \text{(A3)}
\end{aligned}
$$

Conditioning on the class label y, Eqs. (A1) and (A3) can be combined into a single covariance matrix, where the (i,j)th element is

$$
\begin{aligned}
\Sigma_{P_z|y}[i,j] &= (\Sigma_{P_x|y}[i,j]+\mu_{P_x|y}[i]\mu_{P_x|y}[j])(\Sigma_{P_h}[i,j] \\
&\quad +\mu_{P_h}[i]\mu_{P_h}[j])+(\Sigma_{P_w}[i,j]+\mu_{P_w}[i]\mu_{P_w}[j]) \\
&\quad +(\mu_{P_x|y}[i]\mu_{P_h}[i]\mu_{P_w}[j] \\
&\quad +\mu_{P_x|y}[j]\mu_{P_h}[j]\mu_{P_w}[i])(1+\delta_{ij}) \\
&\quad -\mu_{P_z|y}[i]\mu_{P_z|y}[j] \\
&= \Sigma_{P_x|y}[i,j]\Sigma_{P_h}[i,j]+\Sigma_{P_w}[i,j] \\
&\quad +\mu_{P_x|y}[i]\mu_{P_x|y}[j]\Sigma_{P_h}[i,j] \\
&\quad +\Sigma_{P_x|y}[i,j]\mu_{P_h}[i]\mu_{P_h}[j] \\
&\quad +2\mu_{P_x|y}[i]\mu_{P_h}[i]\mu_{P_w}[i]\delta_{ij} \\
&= \Sigma_{P_x|y}[i,j]\Sigma_{P_h}[i,j]+\sigma_w^4\delta_{ij} \\
&\quad +\mu_{P_x|y}[i]\mu_{P_x|y}[j]\Sigma_{P_h}[i,j] \\
&\quad +\Sigma_{P_x|y}[i,j]\mu_{P_h}[i]\mu_{P_h}[j] \\
&\quad +2\sigma_w^2\mu_{P_x|y}[i]\mu_{P_h}[i]\delta_{ij},
\end{aligned}
$$

where $\delta_{ij}=1$ if $i=j$ or 0 otherwise. The last step in the derivation holds since $E[P_w(f_i)P_w(f_j)]=E[W_iW_i^*W_jW_j^*]$ is $2(\sigma_w^2)^2$ for $i=j$ or is $(\sigma_w^2)^2$ for $i\neq j$ by Isserlis' Gaussian moment theorem,[35] and $\mu_{P_w}[i]=\sigma_w^2$. The result is rewritten in compact form in Eq. (11).

[1] R. L. Field, "Transient signal distortion in a multipath environment," Proc. OCEANS 1990, pp. 111–114.

[2] D. W. Tufts, I. Kirsteins, R. J. Vaccaro, C. S. Ramalingam, and A. Shah, "Improvements in signal processing for time-varying multipath propaga-

tion environments," Proc. OCEANS, 1991, Vol. **1**, pp. 586–593.
[3]S. Senmoto and D. Childers, "Signal resolution via digital inverse filtering," IEEE Trans. Aerosp. Electron. Syst. **8**, 644–640 (1972).
[4]J. E. Ehrenberg, T. E. Ewart, and R. D. Morris, "Signal-processing techniques for resolving individual pulses in a multipath signal," J. Acoust. Soc. Am. **63**, 1861–1865 (1978).
[5]T. E. Ewart, J. E. Ehrenberg, and S. A. Reynolds, "Observations of the phase and amplitude of individual Fermat paths in a multipath environment," J. Acoust. Soc. Am. **63**, 1801–1808 (1978).
[6]F. B. Shin, D. H. Kil, and R. Wayland, "IER clutter reduction in shallow water," Proc. IEEE ICASSP, 1996, pp. 3149–3152.
[7]D. J. Strausberger, E. D. Garber, N. F. Chamberlain, and E. K. Walton, "Modeling and performance of HF/OTH radar target classification systems," IEEE Trans. Aerosp. Electron. Syst. **28**, 396–403 (1992).
[8]G. Okopal and P. Loughlin, "Feature extraction for classification of signals propagating in channels with dispersion and dissipation," Proc. SPIE **6566**, April 9–13, 2007.
[9]H. Liu, P. Runkle, and L. Carin, "Classification of distant targets situated near channel bottoms," J. Acoust. Soc. Am. **115**, 1185–1197 (2004).
[10]D. A. Caughey and R. L. Kirlin, "Blind deconvolution of echosounder envelopes," Proc. IEEE ICASSP, 1996, pp. 3149–3152.
[11]M. J. Roan, M. R. Gramann, J. G. Erling, and L. H. Sibuld, "Blind deconvolution applied to acoustical systems identification with supporting experimental results," J. Acoust. Soc. Am. **114**, 1988–1996 (2003).
[12]D. H. Kil and F. B. Shin, *Pattern Recognition and Prediction with Applications to Signal Characterization* (American Institute of Physics Press, Woodbury, NY, 1996).
[13]M. K. Broadhead and L. A. Pflug, "Performance of some sparseness criterion blind deconvolution methods in the presence of noise," J. Acoust. Soc. Am. **107**, 885–893 (2000).
[14]Q. Zhu and B. Steinberg, "Correction of multipath interference using clean and spatial location diversity," Proc. IEEE Ultrasonics Symp., 1995, pp. 1367–1370.
[15]R. A. Wiggins, "Minimum entropy deconvolution," Geoexploration **16**, 21–35 (1978).
[16]C. A. Cabrelli, "Minimum entropy deconvolution and simplicity: A noniterative algorithm," Geophysics **50**, 394–413 (1984).
[17]M. R. Gupta, H. S. Anderson, and Y. Chen, "Joint deconvolution and classification for signals with multipath," Proc. IEEE ICASSP, 2007.
[18]M. R. Gupta and H. Anderson, "Maximum likelihood signal classification using second-order blind deconvolution probability model," Proc. IEEE Wkshp. Stat. Sig. Proc., 2007.
[19]N. Dasgupta and L. Carin, "Time-reversal imaging for classification of submerged elastic targets via Gibbs sampling and the relevance vector machine," J. Acoust. Soc. Am. **117**, 1999–2011 (2005).
[20]A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice Hall, Englewood Cliffs, NJ, 1989).
[21]A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems* (Springer, New York, 1996).
[22]S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
[23]R. W. Yeung, *A First Course in Information Theory* (Springer, New York, 2002).
[24]E. Y. Lam and J. W. Goodman, "Iterative statistical approach to blind image deconvolution," J. Opt. Soc. Am. A **17**, 1177–1184 (2000).
[25]T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer, New York, 2001).
[26]J. H. Friedman, "Regularized discriminant analysis," J. Am. Stat. Assoc. **84**, 165–175 (1989).
[27]S. Srivastava, M. R. Gupta, and B. A. Frigyik, "Bayesian quadratic discriminant analysis," J. Mach. Learn. Res. **8**, 1287–1314 (2007).
[28]E. Candes, "Compressive sampling," Proc. Int. Congress of Math., 2006, Vol. **3**, pp. 1433–1452.
[29]R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. (Wiley, New York, 2001).
[30]R. P. Goddard, "The Sonar Simulation Toolset," Proc. OCEANS, 1989, Vol. **4**, pp. 1217–1222.
[31]R. P. Goddard, "The Sonar Simulation Toolset, Release 4.1: Science, mathematics and algorithms," Technical Report APL-UW TR 0404, Applied Phyics Laboratory, University of Washington, 2005; URL http://stinet.dtic.mil/cgi-bin/GetTRDoc?AD=ADA430617&Location=U2&doc=GetTRDoc.pdf (Last viewed August 2008).
[32]R. E. Fan, P. H. Chen, and C. J. Lin, "Working set selection using the second order information for training SVM," J. Mach. Learn. Res. **6**, 1889–1918 (2005).
[33]D. K. Mellinger and C. W. Clark, "Mobysound: A reference archive for studying automatic recognition of marine mammal sounds," Appl. Acoust. **67**, 1226–1242 (2006); http://hmsc.oregonstate.edu./projects/MobySound/ (Last viewed: August 2008).
[34]C. Therrien, *Discrete Random Signals and Statistical Signal Processing* (Prentice Hall, Upper Saddle River, NJ, 1992).
[35]L. Isserlis, "On a formula for the product-mean coefficient of any order of a normal frequency distribution in any number of variables," Biometrika **12**, 134–139 (1918).

Transient evoked otoacoustic emission latency and estimates of cochlear tuning in preterm neonates

Arturo Moleti[a)]
Dipartimento di Fisica, Università di Roma "Tor Vergata," Via della Ricerca Scientifica, 1, 00133 Roma, Italy

Renata Sisto[b)]
Dipartimento Igiene del Lavoro, ISPESL, Via Fontana Candida, 1, 00040 Monte Porzio Catone (Roma), Italy

Alessia Paglialonga[c)]
Istituto di Ingegneria Biomedica, CNR, and Dipartimento di Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

Federica Sibella[d)]
Istituto di Ingegneria Biomedica, CNR, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

Lucien Anteunis[e)]
Department of Otorhinolaryngology and Head and Neck Surgery, University Hospital Maastricht, P.O. Box 5800, Maastricht, Limburg 6202 AZ, The Netherlands

Marta Parazzini[f)] and Gabriella Tognola[g)]
Istituto di Ingegneria Biomedica, CNR, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

(Received 23 April 2008; revised 29 July 2008; accepted 29 July 2008)

The latency of transient evoked otoacoustic emissions has been evaluated in a sample of 58 ears from 34 preterm neonates, to understand if the estimates of cochlear tuning based on the otoacoustic emission latency show signs of developmental changes. A previous study on the same otoacoustic emissions analyzed here [Tognola *et al.* (2005). "Cochlear maturation and otoacoustic emissions in preterm infants: A time-frequency approach," Hear. Res., **199**, 71–80] reported indeed a significant change in the otoacoustic emission latency with postconception age. This last result, which would imply a significant decrease of tuning, was partially biased by the presence of spontaneous emissions. In this study, the same neonate data are reanalyzed using a novel time-frequency algorithm, less sensitive to spontaneous emissions. Asymmetry between right and left ears has been found, with the left ears showing no significant change, whereas in the right ears and in the 1.5–2.5 kHz frequency range only, a slow decrease of latency with postconception age (0.1–0.2 ms/week) was observed. The correspondent tuning estimates based on latency decrease by 0.4–0.5/week. Significant differences between neonate and adult latency were confirmed, which could be either cochlear or middle ear in nature. These findings are compared to previous studies on distortion product suppression tuning curves in preterm neonates.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977737]

PACS number(s): 43.64.Jb [BLM] Pages: 2984–2994

I. INTRODUCTION

In this study, transient evoked otoacoustic emissions (TEOAEs) from 58 preterm neonate ears were time-frequency analyzed to measure response latency and to derive estimates of cochlear tuning between 29 and 46 weeks of postconception age (PCA). Latency changes correlated with increasing PCA would be an indication for maturation of the hearing system during this period, centered on the full-term PCA. In neonates, one has obviously to rely on fast and objective techniques that do not require any active collaboration from the subject. Therefore OAEs seem to be a really powerful tool for investigating this particular issue.

OAEs are a by-product of the active cochlear mechanisms and provide important information about the physiology of hearing (Probst *et al.*, 1991). The standard OAE classification is based on the evoking technique: Spontaneous OAEs (SOAEs) are recorded without any stimulus, TEOAEs are the response to broadband transient stimulus, stimulus frequency OAEs (SFOAEs) are evoked by a pure tone, and distortion product OAEs (DPOAEs) are the nonlinear response to two tones of nearby frequencies. A more recent and accurate taxonomy (Shera and Guinan, 1999) classifies

[a)]Electronic mail: arturo.moleti@roma2.infn.it
[b)]Electronic mail: renata.sisto@ispesl.it
[c)]Electronic mail: alessia.paglialonga@polimi.it
[d)]Electronic mail: federica.sibella@polimi.it
[e)]Electronic mail: l.anteunis@skno.azm.nl
[f)]Electronic mail: marta.parazzini@polimi.it
[g)]Electronic mail: gabriella.tognola@polimi.it

 © 2008 Acoustical Society of America

OAEs according to their generation mechanism, distinguishing between OAEs generated by nonlinear distortion (wave-fixed in nature) and those generated by linear reflection from cochlear roughness (place-fixed). The two mechanisms contribute to the TEOAE, DPOAE, and SFOAE responses, and can be identified, in principle, from their different phase/frequency relation.

The TEOAE latency is defined as the time delay between the transient stimulus and the maximum of the evoked OAE amplitude. The acoustic stimulus propagates forward along the basilar membrane (BM) as a slow transverse traveling wave (TW). There is no general consensus about the backward transmission mechanism of OAEs. Transmission-line cochlear models (see, e.g., Furst and Lapid, 1988; Zweig and Shera, 1995; Talmadge *et al.*, 1998; and Shera *et al.*, 2005) predict that, for each frequency component, a slow backward TW packet is generated near the resonant place, transmitted back through the middle ear, and recorded as an OAE in the ear canal. Other cochlear models (Wilson, 1980; Vetesnik *et al.*, 2006) assume that most of the otoacoustic energy propagates back via fast longitudinal compression waves in the fluid. In the last case, the OAE backward transmission would imply negligible time delay. Here, for simplicity, we will not consider this possibility, but we wish to remark that it would imply a significant change in the estimates of cochlear tuning based on the measure of the OAE latency (Moleti and Sisto, 2008). Therefore, we assume that, for each frequency within the stimulus bandwidth, latency consists of the traveling time along the cochlear membrane from the base to its tonotopic place and back (cochlear round-trip delay). As tonotopic places are distributed along the cochlea according to the Greenwood frequency map (Greenwood, 1990), latency is intrinsically frequency dependent. In particular, the relation between latency and frequency is dependent on the width of the cochlear filter through the dependence of the traveling wave velocity (TWV) on the sharpness of the resonance (see, e.g., Sisto and Moleti, 2007). Thus, measurements of the TEOAE latency may provide an indirect estimate of the tuning curve, which could be particularly useful for subjects that, like neonates, cannot be studied with psychoacoustical techniques. More generally, as with other OAE measurable parameters (DPOAE phase-gradient delays, TEOAE stimulus-response curves, TEOAE decay time, etc.), TEOAE latency provides important information about the cochlear transmission and amplification mechanisms. Cochlear transmission and amplification are related because both the amplifier gain and the TWV are dependent on the sharpness of the cochlear resonances, expressed by the quality factor Q, defined as the ratio between frequency and bandwidth.

Several time-frequency analysis techniques are available, based, e.g., on the wavelet transform (WT) (Wit *et al.*, 1994; Tognola *et al.*, 1997; Sisto and Moleti, 2007) or on matching pursuit (MP) algorithms (Jedrzejczak *et al.*, 2004; Notaro *et al.*, 2007), which are all capable of providing reliable estimates of the relation between TEOAE latency and frequency. MP analysis is a hierarchical recursive procedure that decomposes the original TEOAE waveform into the sum of "atoms," which are wave packets of given amplitude, latency, frequency, and time span (inverse of the atom bandwidth). It can be useful to identify atoms or groups of atoms associated with peculiar cochlear phenomena, such as multiple internal reflections, intermodulation distortion products, and SOAEs, whose latency-frequency relation significantly deviates from the norm (Jedrzejczak *et al.*, 2007; Notaro *et al.*, 2007).

Cochlear latency is also related to the $2f_1$-f_2 DPOAE group delay (more correctly named phase-gradient delay) (Moulin and Kemp, 1996a; Prijs *et al.*, 2000; Schoonhoven *et al.*, 2001). The DPOAE phase-gradient delay is a characteristic time associated with the slope of the functional relation between the DPOAE phase and the DPOAE frequency $f_{DP}=2f_1-f_2$, related to the round-trip delay introduced by the forward transmission along the cochlear membrane of the primary tones from the source to the DPOAE generation place and the backward transmission of the DPOAE to the microphone. The interpretations of the DPOAE phase-gradient delay data in terms of cochlear tuning (Moulin and Kemp, 1996b; Bowman *et al.*, 1998) have been questioned (Shera *et al.*, 2000), mainly because it is not possible to disentangle the transmission delay from the filter buildup time. Shera *et al.* (2002) measured SFOAE group delays in humans and small mammals, which also provides an interpretation of the results in terms of cochlear tuning (Shera and Guinan, 2003). A recent study (Sisto and Moleti, 2007) provided accurate measurements of the TEOAE latency and model-dependent estimates of cochlear tuning at different click stimulus levels, between 60 and 90 dB pSPL. The decrease in latency with stimulus level was interpreted as due to reduced cochlear tuning at high levels of stimulation.

A few TEOAE latency studies focused on neonates, showing that (1) the neonate latency is significantly longer than that of adults (Sisto and Moleti, 2002); (2) the latency is longer in "fail" neonates, compared to "pass" neonates (Moleti *et al.*, 2005); and (3) the latency of premature neonates is longer than that of term neonates and decreases with increasing PCA (Tognola *et al.*, 2005). This last result was partially biased by the presence of SOAEs and by the definition of latency (time necessary to collect 50% of the total band energy) adopted in that study. As a matter of fact, Jedrzejczak *et al.* (2007), comparing the latency of preterm and full-term neonates, found that the latency difference between them was strongly reduced when the effect of SOAEs was taken into account. They identified in premature neonates a large number of long-lasting MP 'atoms', associated with SOAEs, whose latency was around the center of the acquisition interval ($\approx$10 ms) and was not correlated with frequency. After having removed these atoms from the analysis, the latency difference between preterm (PCA=33 weeks) and full-term neonates was significant only in one frequency band (2.8 kHz), whereas, including all atoms, significant differences were observed at 1.4, 2, and 2.8 kHz (but not at 4 kHz). They also found that the number of these long-lasting atoms per ear was higher at a lower PCA, although they acknowledged that there was no previous evidence in literature for prevalence of SOAEs in preterm in comparison to full-term neonates (Morlet *et al.*, 1995). At high frequency, the typical TEOAE latency is much shorter than half

the acquisition time; therefore, even a small fraction of ears with SOAEs can produce significant overestimate of the average latency of the sample.

The age dependence of the cochlear tuning estimated from TEOAE latency may be compared to that of other tuning estimates, based on the width of the DPOAE suppression tuning curves (STCs). No significant decrease in tuning with PCA has been found in neonates, both in cross-sectional and longitudinal studies (Abdala, 2003, 2004; Abdala *et al.*, 2007). These studies show that there is no significant change of cochlear tuning from PCA=33 weeks to 6 months of age, whereas there is a significant tuning difference between six month infants and adults. The last difference is not necessarily cochlear in nature because it could also be explained by maturation of the middle ear forward transmission (Abdala and Keefe, 2006; Abdala *et al.*, 2007). In fact, increased forward transmission in adults would result in higher effective stimulus levels at the cochlea input and, consequently, lower tuning because the cochlear amplifier gain decreases with stimulus level.

The method proposed by Moleti and Sisto (2003) and refined in (Sisto and Moleti, 2007) has been used to estimate cochlear tuning from TEOAE latency measurements. The purpose of this study is to understand if the estimates of cochlear tuning based on the measure of the TEOAE latency show any significant developmental change. A previous study on the same click evoked OAEs analyzed here (Tognola *et al.*, 2005) reported a significant change with PCA of TEOAE level and latency. The reported change of latency would imply, according to transmission line cochlear models, a very significant decrease in tuning with PCA. In this study, we go into further details in the analysis of TEOAE latency and tuning of the same preterm neonates by developing a novel algorithm, less sensitive to the presence of spontaneous emissions, to derive the latency (and tuning) of the TEOAE frequency components. The goal is to understand what fraction of the previously observed dependence of latency on PCA is not attributable to the presence of SOAEs. Also, comparisons with studies of the DPOAE STCs will be done.

II. MATERIALS AND METHODS

A. TEOAE recordings

TEOAEs were measured from 58 ears of 34 preterm neonates (24 out of 34 had bilateral measurements and 10 had unilateral measurements), which were admitted to the Neonatal Intensive Care Unit. Their mean gestational age at birth was 30.3 weeks (SD=1.4 weeks). All preterm neonates showed one or more of the following risk factors for sensorineural hearing impairment (Joint Committee on Infant Hearing, 1991): very low birth weight (<1500 g), administration of potential ototoxic drugs (for some preterm infants for more than 5 days), depression (low APGAR score (Apgar, 1953) or asphyxia) at birth, and prolonged mechanical ventilation for more than 9 days. None of the risk factors occurred during the period of TEOAE measurement.

Each ear had a minimum of 3 and a maximum of 11 TEOAE measurements in a period ranging from 1 to 8 weeks after birth. A total of 307 TEOAEs were thus obtained. For both preterm and full-term newborns, an otoscopic examination was performed before each OAE measurement to ensure the absence of middle ear effusion.

The preterm neonate data analyzed here are the same as those analyzed in a previously published study by some of the current authors; thus, further details on preterm neonates can be found in (Tognola *et al.*, 2005). In addition to preterm infants, TEOAEs from 333 pass ears of full-term neonates and from 94 normal-hearing young adults (age =18–25 years) were also considered for comparison, already analyzed, respectively, in (Moleti *et al.*, 2005) and (Sisto and Moleti, 2002). TEOAEs from full-term babies were recorded at the third day after delivery during a universal hearing screening program at the Service of Neurophysiopathology of the Mangiagalli Clinic (Milan, Italy). For both term-born neonate and adult TEOAEs considered here, an emission was defined as pass (Vohr and Maxon, 1996) when the overall reproducibility (i.e., the zero-lag cross-correlation coefficient between the *A* and *B* replicates in the 2.5–20 ms window) was $\geq$0.7 and the reproducibilities in 0.8-kHz-wide frequency bands were $\geq$0.5 at 1.6 kHz and $\geq$0.7 at 2.4, 3.2, and 4.0 kHz. All TEOAEs were recorded using an ILO88 system (Otodynamics, Ltd., Hatfield, UK), digitized at a rate of 25 000 samples/s. For each ear, two sets of replicate responses (*A* and *B* replicates recordings in ILO equipment) to 260 repetitions of the click-train (four clicks per train) were averaged, according to the "nonlinear" mode of operation (Bray and Kemp, 1987; Kemp *et al.*, 1986). The averaged data were windowed and digitally filtered using the default ILO window (2.5–20 ms, cosine-tapered from 2.5 to 5 ms) and filter (second order digital band pass set at 600–6000 Hz). All the responses included in the examined dataset were checked to have stimulus stability greater than 75% and stimulus level ranging from 77 to 83 dB peak.

B. Time-frequency analysis

The 20 ms TEOAE waveforms were analyzed by a time-frequency WT technique (for more details, see Tognola *et al.*, 1997, 1998). The WT at the generic time t and frequency f of signal $x(t)$ is defined as (Mallat, 1989)

$$\mathrm{WT}(\tau,f)=\int_t x(t)\left(\frac{f}{f_0}\right)^{-2}\sqrt{\frac{f}{f_0}}\psi^*\left(\frac{f}{f_0}(t-\tau)\right)dt, \qquad (1)$$

where * denotes complex conjugation. Basically, the WT decomposes a signal $x(t)$ into elementary components by means of a bank of band-pass filters, $\sqrt{f/f_0}\psi(f/f_0(t))$, that are iteratively derived from a prototype function $\psi(t)$, which is called the mother wavelet. The mother wavelet $\psi(t)$ is an oscillating function with zero mean and finite energy, centered on time $t=0$; its Fourier transform is a band-pass function centered on frequency f_0. The bandwidth of the band-pass filters derived from the mother wavelet is proportional to the central frequency of the filter, whereas the duration of their impulse response in the time domain is inversely proportional to the central frequency of the filter (for a more detailed mathematical treatment, see Tognola *et al.*, 1997, 1998). Roughly speaking, this type of analysis provides a

decomposition of a signal into subbands with a bandwidth linearly increasing with frequency.

The root-mean-squared (rms) value of coefficients $WT(\tau,f)$ was computed over 500-Hz-wide frequency intervals to get a set of WT band coefficients $\overline{WT(\tau,f_i)}$ as follows:

$$\overline{WT(\tau,f_i)} = \sqrt{\frac{1}{\Delta f}\int_{f_i-0.5\Delta f}^{f_i+0.5\Delta f} WT^2(\tau,f)df}, \quad (2)$$

where f_i $(i=1,\ldots,7)$ is equal to 1, 1.5, 2, 2.5, 3, 3.5, and 4 kHz and $\Delta f=500$ Hz. The set of WT band coefficients $\overline{WT(\tau,f_i)}$ at a given frequency f_i represents the time evolution of the signal within the spectral band centered in f_i.

C. Latency definition

For each ear j, the latency $\tau_j(f_i)$ of the WT coefficients of the band centered in f_i was defined in this study as the time at which the envelope of the function $\overline{WT_j(\tau,f_i)}$ reaches its maximum value. Due to oscillations of the wavelet coefficients, it is better to evaluate the maximum of their envelope, either by performing a polynomial fit or any other kind of smoothing. In this study, the maximum of the envelope was evaluated as the average time of the points of the function $\overline{WT_j(\tau,f_i)}$ exceeding a given threshold, fixed at 95% of its maximum. By doing so, one gets a statistical distribution of the individual band latency $\tau_j(f_i)$ for each ear and each frequency. Individual latencies $\tau_j(f_i)$ at a given frequency band f_i were then pooled together by computing the average value across ears for each PCA group; this average value was called $\tau^*(f_i)$. Before computing the average, the individual band latencies $\tau_j(f_i)$ from bands with signal-to-noise ratio (SNR) lower than unity were discarded, to avoid introducing a bias associated with the presence of noisy data, which tend to give latency values randomly distributed within the acquisition interval. At high frequency, this implies an overestimate of the average latency that is much relevant for noisier records (Sisto and Moleti, 2007). If the SNR increases with PCA (Tognola *et al.*, 2005), this effect may simulate a development of the cochlear function.

As an example, the top panel of Fig. 1 shows the envelopes of the function $\overline{WT_j(\tau,f_i)}$ of the frequency band $f_i=4$ kHz, extracted from the TEOAEs of subject P011 measured at four different PCAs. It may be noted that, at this frequency, as PCA increases there is a slow decrease in the peak time, which, according to our definition of latency $\tau_j(f_i)$, means that latency decreases, and there is also an evident decrease in the height of the tail extending to long latency values. This tail is typically associated with the contributions from SOAEs and long-lasting emissions, whose number has been reported to decrease with PCA (Jedrzejczak *et al.*, 2007). The definition of latency used in the previous paper (Tognola *et al.*, 2005), i.e., latency defined as the time needed to collect 50% of the total band energy, was sensitive to both the position of the peak and the height of the tail; so, this former definition of latency was affected by the presence of SOAEs. Conversely, the latency definition used in this study is mainly sensitive to the position of the peak of the envelope and, thus, more effectively handles the presence of SOAEs.

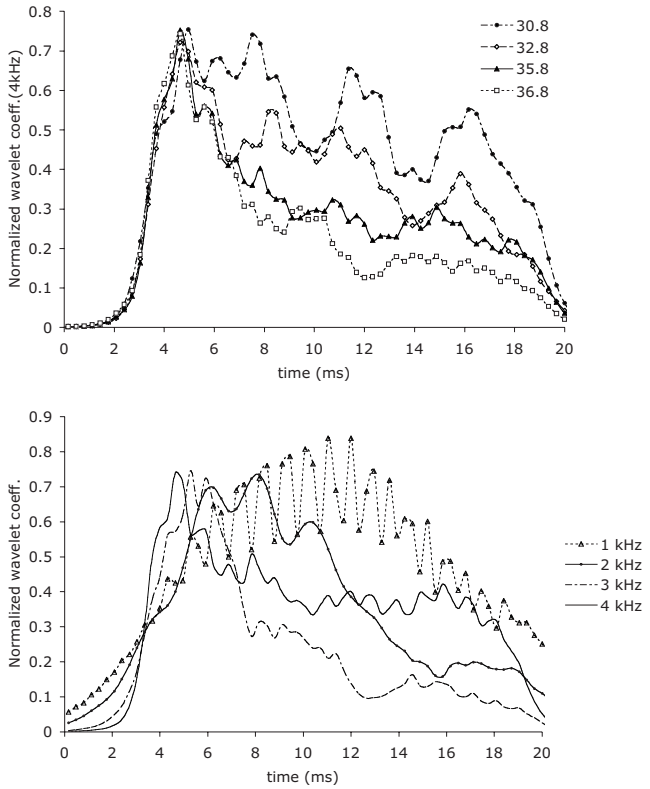


FIG. 1. (Top) Individual wavelet coefficients for subject P011 at different PCAs at 4 kHz. The height of the long-latency tail tends to decrease with increasing PCA, probably because the fraction of energy associated with SOAEs decreases. (Bottom) Wavelet coefficients of the same subject for four frequency bands between 1 and 4 kHz.

The bottom panel of Fig. 1 shows again the envelopes of the function $\overline{WT_j(\tau,f_i)}$ for four frequency bands extracted from the TEOAEs of the previous subject recorded at 35.8 week PCA. Note the low latency cutoff due to the acquisition window, which suggests that one should not fully trust the latency estimates for frequencies higher than 3 kHz. In fact, above 3 kHz, the peak is close to the onset of the acquisition window, so the latency starts to be systematically overestimated.

D. Cochlear tuning estimation

The method described in (Sisto and Moleti, 2007) was applied to the average latency data $\tau^*(f_i)$ to get an estimate of the neonatal cochlear tuning curve. We recall that the measured latency $\tau^*(f_i)$ at a given frequency band centered at frequency f_i is made up of a cochlear contribution, $\tau_c(f_i)$, due to the round-trip delay associated with the transmission of the acoustic signal along the cochlear membrane, and of a much smaller contribution from the probe and the middle ear, τ_{nc}, (of order 0.1–0.2 ms, as estimated from the slope of the phase of the middle ear round-trip transfer function (Puria, 2003)), which will be neglected. The middle ear transmission can play a significant role in this study by altering

the amplitude of the stimulus reaching the cochlear base in a frequency- and age-dependent way, but it does not directly introduce a significant physical delay.

The method is based on a simple transmission line cochlear model, in which the round-trip cochlear delay $\tau_c(f_i)$ is computed as twice the integral of the inverse TWV from the base to the tonotopic place,

$$\tau_c(\omega_i) = 2\int_0^{x(\omega_i)} dx \mathrm{TWV}^{-1} = 2\int_0^{x(\omega_i)} dx \frac{\partial k}{\partial \omega}. \quad (3)$$

The TWV is a function of the relation between the wave number k, angular frequency $\omega_i = 2\pi f_i$, and cochlear position x, and drops to a minimum near the resonant place. As this slowing down is a function of the resonance quality factor Q, it is possible to get an estimate of the tuning curve from the measured latency $\tau^*(f_i)$ as follows:

$$Q(\omega_i) \approx \frac{1}{\beta^2}\left(\frac{\omega k_\omega}{k_0}(\tau^*(\omega_i) - \tau_{nc}) + \frac{2\omega_i}{\omega_{\max}}\right)^2, \quad (4)$$

where $\tau(\omega_i)$ is the experimental average latency value and $\omega_i = 2\pi f_i$. The inverse length scale $k_\omega = 1.382\ \mathrm{cm}^{-1}$ and the maximum frequency $\omega_{\max} = 2\pi 20.655$ krad/s are the two parameters of a simplified Greenwood map as follows:

$$\omega(x) \approx \omega_{\max} e^{-k_\omega x}, \quad (5)$$

in which the scale-invariance breaking additive constant term has been neglected. The constant $k_0 = 31\ \mathrm{cm}^{-1}$ is associated with the geometry and density of the BM and of the cochlear cavity (Talmadge *et al.*, 1998), and $\beta \approx 1.2$ is a dimensionless model-dependent constant (for more details about the model, see Moleti and Sisto, 2003 and Sisto and Moleti, 2007).

The experimental TEOAE latency data are in good agreement with a simple power law

$$\tau(f) = af^b, \quad (6)$$

with f in kHz. According to transmission-line cochlear models, the numerical value of b is related to the relation between cochlear tuning and frequency, and scale-invariant models, for which Q has to be constant, predict that b should be exactly equal to -1 (Talmadge *et al.*, 1998). The experimental estimates are consistent with a of order 10 ms and b in the range -0.4 to -0.7 (Tognola *et al.*, 1997; Sisto and Moleti, 2002; 2007). The fact that the experimental data yield b significantly different from the scale-invariant prediction reflects the significant violation of the scaling symmetry associated with the increase in cochlear tuning with frequency (Sisto and Moleti, 2007).

III. RESULTS AND DISCUSSION

A. Cross-sectional analysis

1. Latency and tuning in seven PCA groups

Here we show the results of a cross-sectional (i.e., inter-subject) analysis, in which an average is performed on latencies of TEOAE bands grouped together according to PCA at the time of the recording. First we considered the following seven PCA groups: <30, 30–32, 32–34, 34–36, 36–38, 38–40, and >40 weeks, with no distinction between left and right ears, as in (Tognola *et al.*, 2005). In the top panel of Fig. 2 we show the relation between average band latency $\tau^*(f_i)$ and frequency for the seven groups of increasing PCA. Average latencies of preterm neonates are also compared to that of term-born neonates and young adults. At all PCAs, latency shows the typical monotonic decrease with frequency. At some frequencies, it could also be barely noticed that the preterm neonate latency tends to slowly decrease with PCA. In (Tognola *et al.*, 2005), this latency decrease with PCA was much more evident, as can be appreciated from a visual comparison with their Fig. 8. In the latency from the present analysis, some decreasing trend is more clearly visible in the bottom panel of Fig. 2, where latency is multiplied by the square root of frequency (in kHz), approximately compensating for its intrinsic frequency dependence, to expand the vertical scale. In any case, latencies of TEOAEs from both preterm and term-born neonates appear significantly longer than those measured from adults, confirming the results of Moleti and Sisto (2003). If we pool together all neonates (term and preterm) and compare their latency distribution with that of adults, both a student's t-test and a Wilcoxon rank sum test show indeed that the latency

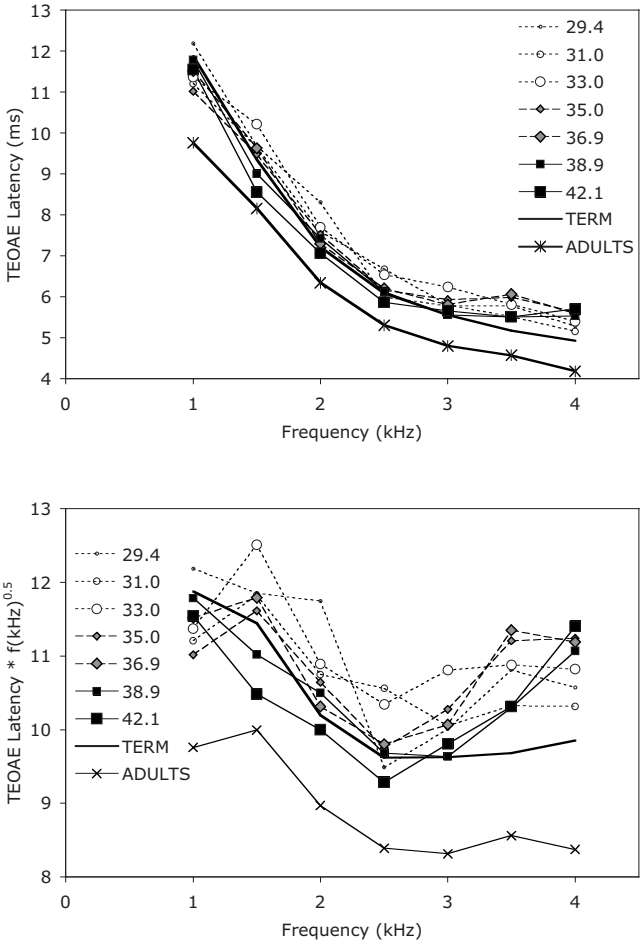


FIG. 2. (Top) Average latency-frequency relation for the seven groups of preterm neonates of increasing PCA. The term group consists of 333 pass full-term neonate ears (with average PCA=39 weeks), whereas the adult group consists of 94 ears of normal-hearing young adults. (Bottom) Same as in the top panel, but latency is multiplied by $f^{0.5}$, to expand the vertical scale, so that slow dependence on PCA may be better appreciated.

TABLE I. Tuning estimates for the seven PCA groups. Above 3 kHz (italic), systematic errors associated with the window onset time could lead to overestimates.

Frequency PCA	1 kHz	1.5 kHz	2 kHz	2.5 kHz	3 kHz	*3.5* kHz	*4* kHz
<30	8.55	12.3	16.3	13.7	18.4	*25*	*27.6*
30–32	7.27	12.3	13.8	16.8	18.5	*22.9*	*26.4*
32–34	7.48	13.7	14.1	16.1	21.2	*25.3*	*28.9*
34–36	7.03	11.8	13.5	14.4	19.3	*26.8*	*31*
36–38	7.66	12.2	12.7	14.6	18.6	*27.4*	*30.7*
38–40	8.01	10.7	13.1	14.2	17.1	*22.8*	*30.1*
>40	7.7	9.71	12	13.1	17.7	*22.9*	*31.8*

difference between neonates and adults is statistically significant, with $p<10^{-6}$, in all the frequency bands from 1 to 4 kHz.

The change of latency and cochlear tuning with PCA in preterm neonates can be better appreciated from Table I and Fig. 3, where the average latency $\tau^*(f_i)$ and quality factor Q (i.e., the cochlear tuning) are plotted as a function of PCA, for frequency bands ranging from 1 to 3 kHz. Error bars represent the standard errors of latency and Q distributions of each PCA group. For the three frequency bands between 1.5 and 2.5 kHz, latency and Q seem to slowly decrease with PCA, but the small number of ears in the first and last PCA groups causes large fluctuations. This decrease in latency with PCA may be fitted to straight lines with slopes of order −0.1 ms/week, while a linear fit to the decrease of Q with PCA yields negative slopes of order −0.3/week. In these frequency bands the latency estimates are most reliable because the typical latency values are comfortably longer than the acquisition window onset time and the TEOAE SNR is typically high. In addition to data from preterm neonates, Fig. 3 shows, for comparison, latency and tuning values of term-born neonates and adults to get a visual impression of the size of the decreasing trend with PCA, if compared to that of the well-established difference between adults and neonates (Moleti and Sisto, 2003). It may be noted that the latency (and the tuning estimates) of the term-born neonates, whose average PCA was 39 weeks, fits the trend of the data of preterm neonates. In other words, at the same PCA, latency of term and preterm neonates is not different, suggesting that premature birth itself does not affect significantly cochlear latency (and tuning).

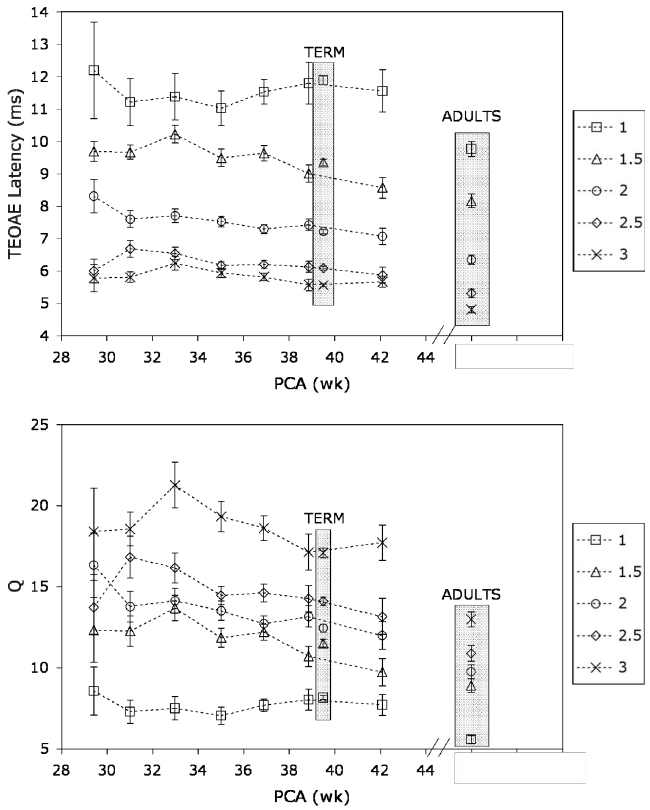


FIG. 3. (Top) Average latency between 1 and 3 kHz for the seven groups defined according to the PCA at the time of the measurement. The horizontal axis represents the average PCA of the sample. (Bottom) Dependence of tuning on PCA for frequencies between 1 and 3 kHz. Cochlear tuning is estimated from Eq. (5) using the definition of latency τ^*. In both panels the term and adult groups are also shown for comparison. The term data points are slightly shifted to the right (by 0.5 weeks) to avoid confusion. Error bars represent standard errors.

2. Dependence of latency and tuning on PCA

The statistical significance of the dependence of latency and tuning on PCA can be quantitatively evaluated by linear regression, without using the seven PCA groups. As the actually measured quantity is the latency, but we focused on the decrease in tuning, we have chosen to apply the analysis both to a linear relation between latency and PCA and to another one between tuning and PCA. Despite the expected quadratic dependence of tuning on latency, as the relative change of latency with PCA is small, of order $\varepsilon \ll 1$, the corresponding quadratic term would be of order ε^2; thus, it is not contradictory to check for linear dependence on PCA for both quantities.

In Fig. 4(a) we show the linear regressions computed on the latency as a function of PCA for frequencies band centered at 1.5 (top row), 2.0 (middle row), and 2.5 kHz (bottom row). The white and black symbols are used for the left and right ears, plotted separately in the two columns of Fig. 4. Similarly to Fig. 4(a) and 4(b) shows linear regressions fitted to the dependence on PCA of the Q values. The evident right-left asymmetry suggested performing separate regressions for the left and right ears, but we report the results also for a regression made on all the ears, for comparison with previous studies.

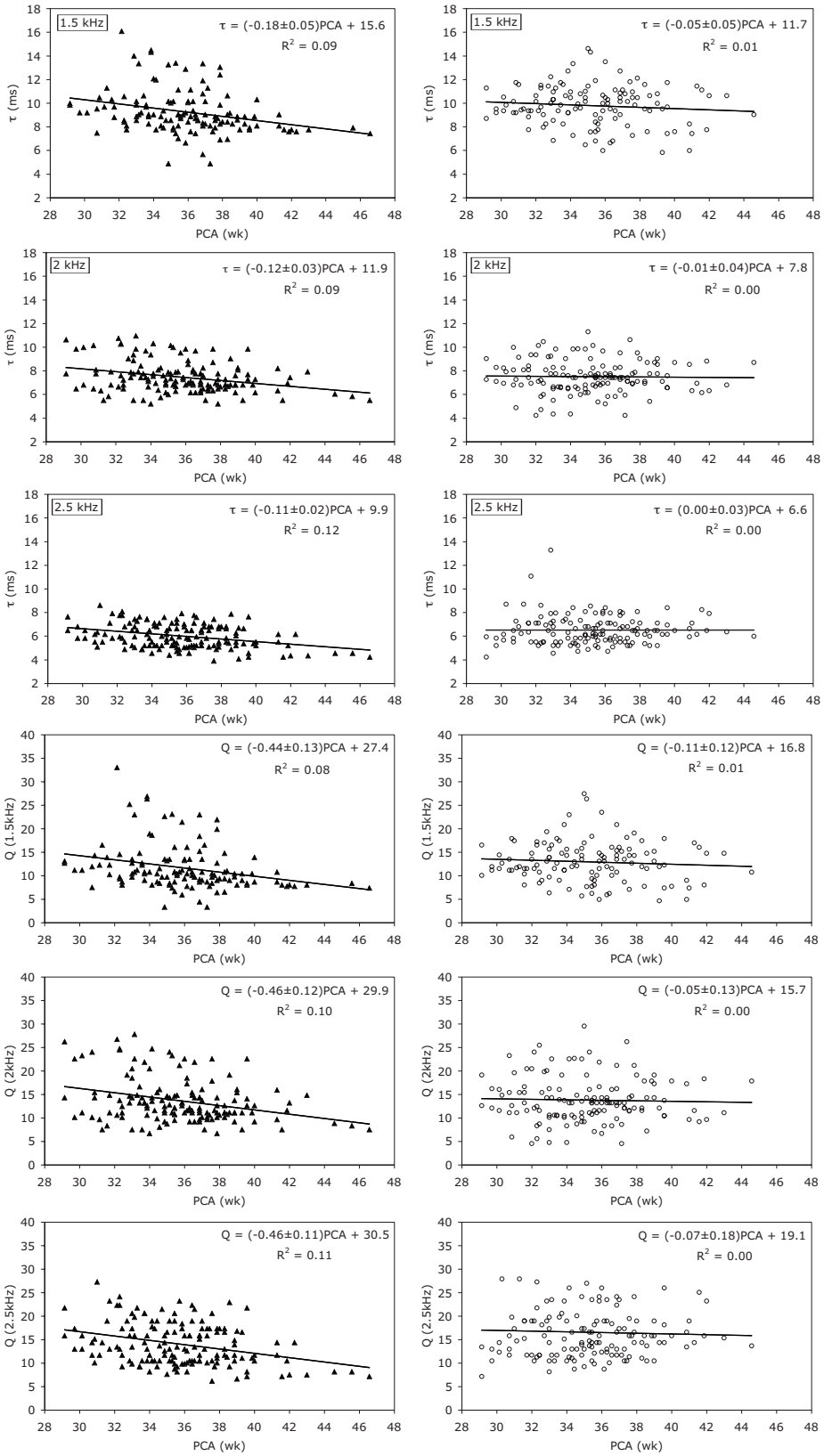


FIG. 4. Individual latency (top) and tuning (bottom) dependence on PCA in the right (first column) and left (second column) ears at 1.5, 2, and 2.5 kHz. Outside this frequency range no significant change with PCA has been observed.

The low R^2 values are due to the large intersubject variability of the OAE latency and show that the statistical association between latency and PCA is rather weak, if many different subjects are analyzed at different ages. In Table II the slopes B and BQ of the linear regressions on latency versus PCA and Q factor versus PCA of all ears are reported together with the statistical errors ΔB and ΔBQ, the value F of the F-distribution, and the correspondent probability p of getting by chance the observed experimental decreasing trend. Despite the low R^2 values shown in Fig. 4, the data of Table II, where right and left ears are pooled and analyzed together, to allow comparisons with previous studies, revealed a statistically significant ($p<0.05$) decrease in latency and tuning with PCA at frequencies of 1.5, 2.0, and

TABLE II. Statistical analysis of the significance of the dependence on PCA of latency and tuning. All ears. The parameters B and BQ are the slopes of the two linear regressions for latency and tuning vs PCA, and ΔB and ΔBQ the corresponding errors. The p-value is associated with the value of F. Statistically significant p-values are printed in bold.

Frequency	1 kHz	1.5 kHz	2 kHz	2.5 kHz	3 kHz	*3.5* kHz	*4* kHz
B (ms/week)	0.00	−0.12	−0.07	−0.07	−0.04	*−0.01*	*0.04*
ΔB	0.08	0.04	0.03	0.02	0.02	*0.03*	*0.03*
F	0.002	11.76	8.06	9.57	3.54	*0.05*	*1.48*
p	0.96	**0.001**	**0.005**	**0.002**	0.06	*0.82*	*0.22*
Frequency	1 kHz	1.5 kHz	2 kHz	2.5 kHz	3 kHz	*3.5* kHz	*4* kHz
BQ (week^{-1})	−0.01	−0.29	−0.27	−0.31	−0.31	*0.04*	*0.31*
ΔBQ	0.10	0.09	0.09	0.11	0.16	*0.36*	*0.47*
F	0.007	10.06	9.29	8.61	3.65	*0.01*	*0.43*
p	0.93	**0.002**	**0.003**	**0.004**	0.06	*0.92*	*0.51*

2.5 kHz. As tuning is directly derived from latency by using Eq. (5), the significance of the decrease in tuning with PCA is obviously related to that of the latency decrease, as shown by the similarity of the p-values of the two analyses. The decrease rate of latency with PCA estimated from TEOAE of all ears ranges from −0.12 ms/week (at 1.5 kHz) to −0.07 ms/week (at 2.5 kHz). Correspondingly, the decrease of Q with PCA ranged from $-0.29/\text{week}^{-1}$ (at 1.5 kHz) to $-0.31/\text{week}^{-1}$ (at 2.5 kHz).

Similarly to Table II, Tables III and IV show statistical data computed separately for right and left ears, respectively. The right-left asymmetry visible in Fig. 4 is quantitatively confirmed, as shown in Tables III and IV: The observed decrease in latency and tuning with PCA is significant ($p < 0.05$) only in the right ears and for the three frequency bands between 1.5 and 2.5 kHz, whereas latency and tuning of left ears do not show any significant change with PCA. For the right ears only, the estimated latency decrease rate was therefore greater than for all ears together and ranged from −0.18 ms/week (at 1.5 kHz) to −0.11 ms/week (at 2.5 kHz), whereas the corresponding decrease of Q ranged from $-0.44/\text{week}^{-1}$ (at 1.5 kHz) to $-0.46/\text{week}^{-1}$ (at 2.5 kHz). The comparison between Tables II and III shows

TABLE III. Statistical analysis of the significance of the dependence on PCA of latency and tuning. Same as Table II, for the right ears only.

Frequency	1 kHz	1.5 kHz	2 kHz	2.5 kHz	3 kHz	*3.5* kHz	*4* kHz
B (ms/week)	−0.09	−0.18	−0.12	−0.11	−0.04	*−0.01*	*0.04*
ΔB	0.11	0.05	0.03	0.02	0.03	*0.04*	*0.04*
F	0.67	11.8	14.8	19.9	1.26	*0.022*	*0.90*
P	0.41	**0.001**	**<0.001**	**<0.001**	0.26	*0.88*	*0.34*
Frequency	1 kHz	1.5 kHz	2 kHz	2.5 kHz	3 kHz	*3.5* kHz	*4* kHz
BQ (week^{-1})	−0.14	−0.44	−0.46	−0.46	−0.30	−0.04	0.32
ΔBQ	0.14	0.13	0.12	0.11	0.24	*0.54*	*0.60*
F	1.09	10.6	15.1	18.2	1.61	*0.005*	*0.29*
P	0.30	**0.001**	**<0.001**	**<0.001**	0.21	*0.94*	*0.59*

TABLE IV. Statistical analysis of the significance of the dependence on PCA of latency and tuning. Same as Table II, for the left ears only.

Frequency	1 kHz	1.5 kHz	2 kHz	2.5 kHz	3 kHz	*3.5* kHz	*4* kHz
B (ms/week)	0.09	−0.05	−0.01	−0.00	−0.05	*−0.01*	*0.02*
ΔB	0.12	0.05	0.04	0.03	0.03	*0.04*	*0.04*
F	0.58	1.19	0.06	0.001	1.98	*0.015*	*0.28*
P	0.45	0.28	0.81	0.98	0.16	*0.90*	*0.59*
Frequency	1 kHz	1.5 kHz	2 kHz	2.5 kHz	3 kHz	*3.5* kHz	*4* kHz
BQ (week^{-1})	0.12	−0.11	−0.05	−0.07	−0.30	*0.04*	*0.18*
ΔBQ	0.15	0.12	0.13	0.18	0.23	*0.49*	*0.73*
F	0.64	0.79	0.16	0.15	1.75	*0.007*	*0.06*
P	0.43	0.38	0.69	0.70	0.19	*0.94*	*0.81*

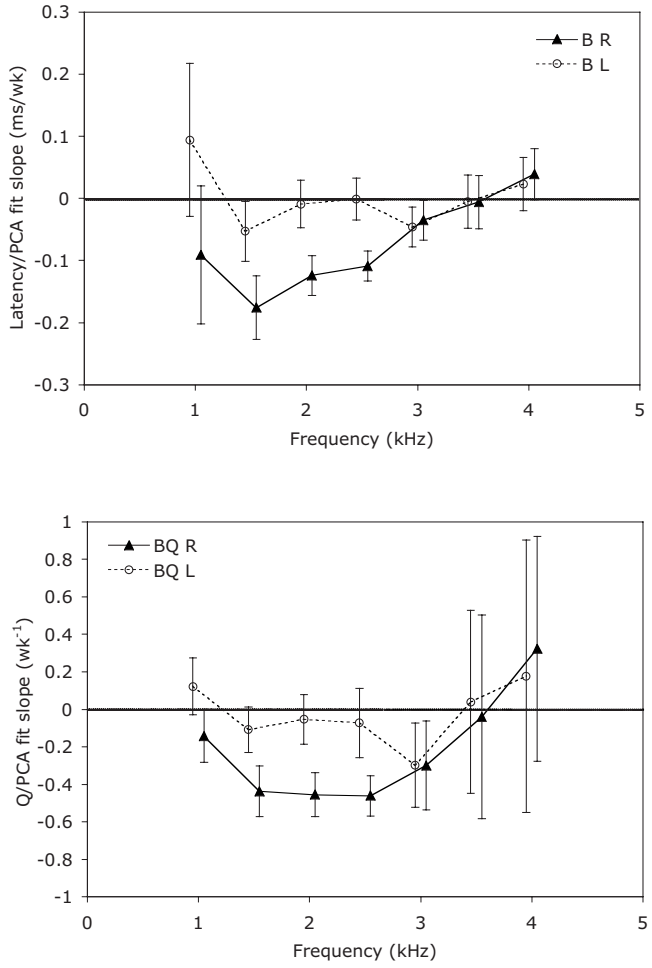


FIG. 5. Slope of the latency/PCA and of the Q/PCA linear regressions between 1 and 4 kHz for the right and left ears, also reported in Tables III and IV.

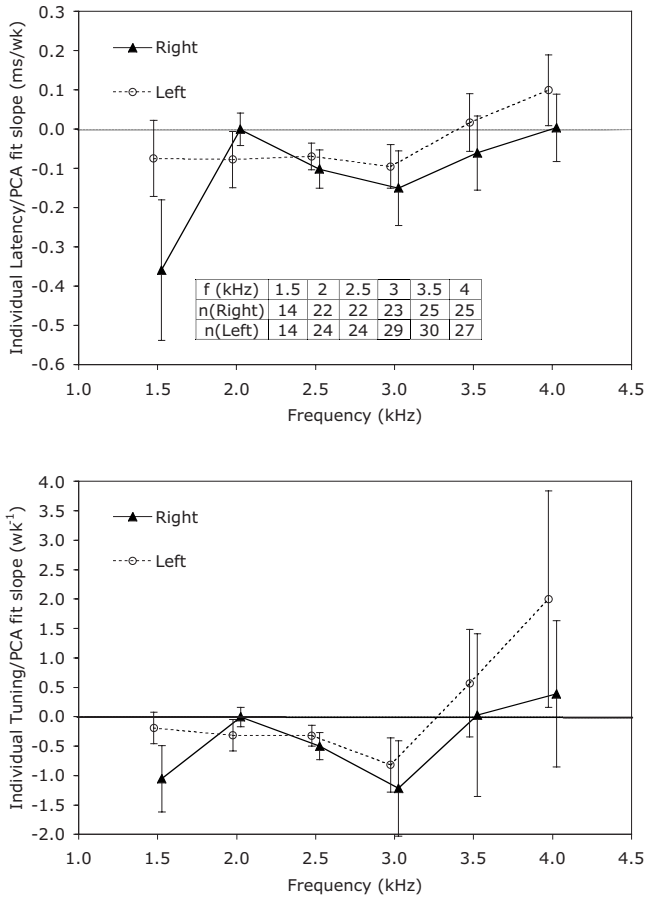


FIG. 6. Longitudinal analysis: individual change of latency (top) and tuning (bottom) with increasing PCA between 1.5 and 4 kHz. The average slope of the individual linear fits is shown for each frequency for the left and right ears. Error bars represent standard errors. The number of ears included in the longitudinal analysis for each frequency band is also reported in the top panel.

that including the left ears in the analysis does not improve the significance of the results, although the size of the sample is doubled. Figure 5 shows the values of the slopes *B* (top panel) and *BQ* (bottom panel) of the linear fits as functions of the frequency (*B* and *BQ* are the same as reported in Tables III and IV).

B. Longitudinal analysis

Parallel to the *cross-sectional* (i.e., intersubject) approach, TEOAEs of the preterm infants were also analyzed with a *longitudinal* (i.e., intrasubject) approach. This was done to check whether the great interindividual variation of latency and *Q* might mask the intraindividual maturation of these particular TEOAE features, when the data are averaged and analyzed cross sectionally. Also, an important feature of the present study is having longitudinally followed each subject during a time interval ranging from 3 to 10 weeks after birth. The small size of the observed evolutionary effect has suggested so far focusing our attention on the cross-sectional analysis to increase the statistical significance of the results. The longitudinal analysis of the change of individual latency/tuning with increasing individual PCA could be important, in principle, to disentangle the effects associated with concurrent phenomena. In fact, any dependence of tuning on PCA could be associated either with physiological development of the hearing function or to some pathological effect of premature birth on tuning, lasting in the following weeks. In our study, the average PCA of the measurements of a given subject was correlated to their PCA at birth, because each subject had been tested only during a period of a few weeks after birth. The longitudinal analysis could in principle provide complementary information to clarify if PCA at birth or PCA at the time of each measurement is the most relevant parameter, but the small size of the observed evolutionary effect, compared with the fluctuations of the individual measurements, has made it impossible to get definitive conclusions.

In the longitudinal analysis of the data we performed, the change of latency with PCA of each of the 58 ears has been obtained by fitting the individual latency versus PCA relation with a linear regression; then, the coefficients representing the slopes of such linear fits were averaged, frequency by frequency, and shown in the top panel of Fig. 6. The average was done separately for left and right ears. For each frequency band, a linear fit was applied only to the ears with SNR >1 within each band at all measurement sessions. In the 1 kHz band, this constraint selected only four ears, so the result for this frequency band is not reported in Fig. 6. Below 3 kHz, a systematically negative slope of latency ver-

sus PCA can be noted, confirming the previous cross-sectional analysis results. Due to the sharp dependence of tuning on latency, a similar analysis applied to the individual dependence of tuning on PCA amplifies the effect of the random fluctuations present in the individual data; thus, the results of the longitudinal analysis are more irregular, particularly at high frequency, as shown in the bottom panel of Fig. 6. Indeed, at high frequency, latency is shorter and close to the window onset time, so systematic errors are more frequent and have a larger effect on the tuning estimates.

IV. DISCUSSION

The results of the cross-sectional and longitudinal analyses show decrease with PCA of the cochlear tuning estimates derived from TEOAE latency measured in preterm neonates with PCA between 29 and 46 weeks. Significant changes are observed only in the right ears and only in the three midfrequency bands (1.5, 2.0, and 2.5 kHz). What was actually measured is a slow decrease in latency with PCA (with slope between −0.11 and −0.18 ms/week) in the right ears, in the three frequency bands between 1.5 and 2.5 kHz. As a consequence, also the corresponding cochlear tuning estimates slowly decrease with PCA (with slope between −0.44 and $-0.46/\text{week}^{-1}$). Latency and cochlear tuning changes with PCA were not significant in the left ears; this asymmetry could either cast some doubts on the accuracy of the study or give some clues for the interpretation of the results. In fact, studies on asymmetry of the peripheral hearing system have shown a higher number and level of SOAEs in the right ears of preterm neonates (Morlet *et al.*, 1995) as well as larger TEOAE amplitudes in the right than in the left ear (Khalfa *et al.*, 1997; 1998). On the other hand, Jedrzejczak *et al.* (2007), who did not perform a separate analysis for right and left ears, have found that the number of long-lasting atoms decreases with increasing PCA. Therefore, a possible explanation of the observed dependence of tuning on PCA (and of its asymmetry) could be related to the presence of long-lasting TEOAE components (frequently associated with SOAEs), whose effect is reduced (with respect to Tognola *et al.*, 2005) but not fully canceled by the different definition of latency used in the present study. It would also be interesting to look for right-left asymmetries in the preterm neonate data analyzed by Jedrzejczak *et al.* (2007).

Results of current study can also be compared with previous results provided by studies based on the measure of DPOAE STCs (Abdala, 2003, 2004), which did not find any significant change of tuning with PCA. A longitudinal study (Abdala, 2003) has shown no significant maturation effect in premature neonates, whereas the premature neonate tuning was found significantly higher than that of adults. A following study (Abdala, 2004), in which STC tuning estimates are reported of subjects ranging from PCA=33 weeks to 3 months of age, showed again no statistically significant difference between term neonates and 3 month infants. Later, this conclusion was extended to six months of age (Abdala *et al.*, 2007). Summarizing, all these DPOAE STC studies show no significant tuning change from PCA=33 weeks to 6 months of age, and a very significant difference between infants and adults. The last result is in agreement with the present study, which also confirms previous findings by Moleti and Sisto (2003). Abdala and Keefe (2006) have suggested that it is not at all sure that the observed difference between adult and infant tuning is really related to maturation of the cochlear amplifier. In fact, due to the relation between cochlear tuning and stimulus level, changes in the forward transmission of the middle ear may result in changes of the effective stimulus level reaching the stapes, and, in this case, the difference in the measured cochlear filter sharpness would be due to the well-known dependence of cochlear tuning on stimulus level. The evolution of STCs and acoustic admittance up to 6 months of age has been compared (Abdala *et al.*, 2007) to confirm this hypothesis. The role of cochlear and middle ear maturation in the evolution of tuning estimates has still to be clarified. These considerations apply also to the results of the present study because it has been demonstrated that the TEOAE latency is a function of stimulus level (Tognola *et al.*, 1997; Sisto and Moleti, 2007). As the stimulus is measured in the ear canal, it is evident that changes with PCA of the input/output matrix of the middle ear can simulate changes of cochlear tuning. If we compare the size of the observed maximum latency changes (2–3 ms) with the dependence on stimulus level of TEOAE latency reported by Sisto and Moleti (2007) in the 1.5–2.5 kHz frequency range, we see that all the observed change could be explained by a 10 dB change in the middle ear forward transmission.

It should be added here that the tuning estimates from DPOAE STCs suffer from uncertainties associated with the complex generation mechanism of DPOAEs, which involves two distinct sources (Shera and Guinan, 1999). The first (nonlinear distortion) source is placed at the cochlear place resonant at frequency f_2, whereas the second (linear reflection) source is close to the f_{DP} place. The typically larger width of DPOAE STCs with respect to behavioral tuning curves is likely related to this fact. The change of DPOAE STCs from birth to adulthood is not necessarily related to overall changes of the cochlear filter but could reflect instead changes in the relative weight of the two DPOAE components. Therefore, our results cannot be directly compared to the above-mentioned DPOAE STC studies also because we measured OAEs in a lower frequency range (1–4 kHz) and at a different stimulus level.

Abdala, C. (**2003**). "A longitudinal study of distortion product otoacoustic emission ipsilateral suppression and input/output characteristics in human neonates," J. Acoust. Soc. Am. **114**, 3239–3250.

Abdala, C. (**2004**). "Distortion product otoacoustic emission (2f1-f2) suppression in 3-month-old infants: Evidence for postnatal maturation of human cochlear function?," J. Acoust. Soc. Am. **116**, 3572–3580.

Abdala, C., and Keefe, D. H. (**2006**). "Effects of middle-ear immaturity on distortion product otoacoustic emission suppression tuning in infant ears," J. Acoust. Soc. Am. **120**, 3832–3842.

Abdala, C., Keefe, D. H., and Oba, S. I. (**2007**). "Distortion product otoacoustic emission suppression tuning and acoustic admittance in human infants: Birth through 6 months," J. Acoust. Soc. Am. **121**, 3617–3627.

Apgar, V. (**1953**). "A proposal for a new method of evaluation of the newborn infant," Curr. Res. Anesth. Analg. **32**, 260–267.

Bowman, D. M., Eggermont, J. J., Brown, D. K., and Kimberley, B. P. (**1998**). "Estimating cochlear filter response properties from distortion product otoacoustic emission (DPOAE) phase delay measurements in normal hearing human adults," Hear. Res. **119**, 14–26.

Bray, P., and Kemp, D. T. (**1987**). "An advanced cochlear echo technique suitable for infant screening," Br. J. Audiol. **21**, 191–204.
Furst, M., and Lapid, M. (**1988**). "A cochlear model for acoustic emissions," J. Acoust. Soc. Am. **84**, 222–229.
Greenwood, D. D. (**1990**). "A cochlear frequency position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (**2004**). "Identification of otoacoustic emissions components by means of adaptive approximations," J. Acoust. Soc. Am. **115**, 2148–2158.
Jedrzejczak, W. W., Hatzopoulos, S., Martini, A., and Blinowska, K. J. (**2007**). "Otoacoustic emissions latency difference between full-term and preterm neonates," Hear. Res. **231**, 54–62.
Joint Committee on Infant Hearing (**1991**). "1990 position statement," AAP News **7**, 6.
Kemp, D. T., Bray, P., Alexander, L., and Brown, A. M. (**1986**). "Acoustic emissions cochleography: Practical aspects," Scand. Audiol. **25**, 71–95.
Khalfa, S., Morlet, T., Micheyl, C., Morgon, A., and Collet, L. (**1997**). "Evidence of peripheral hearing asymmetry in humans: Clinical implications," Acta Oto-Laryngol. **117**, 192–196.
Khalfa, S., Micheyl, C., Veuillet, E., and Collet, L. (**1998**). "Peripheral auditory lateralization assessment using TEOAEs," Hear. Res. **121**, 29–34.
Mallat, S. G. (**1989**). "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693.
Moleti, A., and Sisto, R. (**2003**). "Objective estimates of cochlear tuning by otoacoustic emission analysis," J. Acoust. Soc. Am. **113**, 423–429.
Moleti, A., Sisto, R., Tognola, G., Parazzini, M., Ravazzani, P., and Grandori, F. (**2005**). "Otoacoustic emission latency, cochlear tuning, and hearing functionality in neonates," J. Acoust. Soc. Am. **118**, 1576–1584.
Moleti, A., and Sisto, R. (**2008**). "Comparison between otoacoustic and auditory brainstem response latencies supports slow backward propagation of otoacoustic emissions," J. Acoust. Soc. Am. **123**, 1495–1503.
Morlet, T., Lapillonne, A., Ferber, C., Duclaus, R., Sann, L., Putet, G., Salle, B., and Collet, L. (**1995**). "Spontaneous otoacoustic emissions in preterm neonates: Prevalence and gender effects," Hear. Res. **90**, 44–54.
Moulin, A., and Kemp, D. T. (**1996a**). "Multicomponent acoustic distortion product otoacoustic emission phase in humans. I. General Characteristics," J. Acoust. Soc. Am. **100**, 1617–1639.
Moulin, A., and Kemp, D. T. (**1996b**). "Multicomponent acoustic distortion product otoacoustic emission phase in humans. II. Implications for distortion product otoacoustic emissions generation," J. Acoust. Soc. Am. **100**, 1640–1662.
Notaro, G., Al-Maamury, A. M., Moleti, A., and Sisto, R. (**2007**). "Wavelet and matching pursuit estimates of the transient-evoked otoacoustic emission latency," J. Acoust. Soc. Am. **122**, 3576–3585.
Prijs, V. F., Schneider, S., and Schoonhoven, R. (**2000**). "Group delays of distortion product otoacoustic emissions: Relating delays measured with F1- and F2-sweep paradigms," J. Acoust. Soc. Am. **107**, 3298–3307.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.
Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.
Schoonhoven, R., Prijs, V. F., and Schneider, S. (**2001**). "DPOAE group delays versus electrophysiological measures of cochlear delay in normal human ears," J. Acoust. Soc. Am. **109**, 1503–1512.
Shera, C. A., and Guinan, J. J. Jr. (**1999**). "Evoked otoacoustic emissions arise from two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., Talmadge, C. L., and Tubis, A. (**2000**). "Interrelations among distortion-product phase-gradient delays: Their connection to scaling symmetry and its breaking," J. Acoust. Soc. Am. **108**, 2933–2948.
Shera, C. A., Guinan, J. J., Jr., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Shera, C. A., and Guinan, J. J., Jr. (**2003**). "Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–2772.
Shera, C. A., Tubis, A., and Talmadge, C. L. (**2005**). "Coherent reflection in a two-dimensional cochlea: Short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," J. Acoust. Soc. Am. **118**, 287–313.
Sisto, R., and Moleti, A. (**2002**). "On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears," J. Acoust. Soc. Am. **111**, 297–308.
Sisto, R., and Moleti, A. (**2007**). "Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels," J. Acoust. Soc. Am. **122**, 2183–2190.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.
Tognola, G., Ravazzani, P., and Grandori, F. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.
Tognola, G., Grandori, F., and Ravazzani, P. (**1998**). "Wavelet analysis of click-evoked otoacoustic emissions," IEEE Trans. Biomed. Eng. **45**, 686–697.
Tognola, G., Parazzini, M., de Jager, P., Brienesse, P., Ravazzani, P., and Grandori, F. (**2005**). "Cochlear maturation and otoacoustic emissions in preterm infants: A time-frequency approach," Hear. Res. **199**, 71–80.
Vetesnik, A., Nobili, R., and Gummer, A. (**2006**). "How does the inner ear generate distortion product otoacoustic emissions? Results from a realistic model of the human cochlea," ORL **68**, 347–352.
Vohr, B. R., and Maxon, A. B. (**1996**). "Screening infants for hearing impairment," J. Pediatr. (St. Louis) **128**, 710–714.
Wilson, J. P. (**1980**). "Model for cochlear echoes and tinnitus based on an observed electrical correlate," Hear. Res. **2**, 527–532.
Wit, H. P., van Dijk, P., and Avan, P. (**1994**). "Wavelet analysis of real ear and synthesized click evoked otoacoustic emissions," Hear. Res. **73**, 141–147.
Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Transient evoked otoacoustic emission input/output function and cochlear reflectivity: Experiment and model

Renata Sisto[a)]
Dipartimento Igiene del Lavoro, ISPESL, Via Fontana Candida, 1, 00040 Monte Porzio Catone (Roma), Italy

Arturo Moleti[b)]
Dipartimento di Fisica, Università di Roma "Tor Vergata," Via della Ricerca Scientifica, 1, 00133 Roma, Italy

(Received 23 June 2008; revised 27 August 2008; accepted 29 August 2008)

The complex input/output function of transient evoked otoacoustic emissions is evaluated at different stimulus levels. The experimental response functions were best fitted to the reflectivity functions predicted by theoretical one-dimensional transmission-line models in the perturbative limit. Along with the otoacoustic emission sources usually considered, linear reflection from roughness (place-fixed) and nonlinear distortion (wave-fixed), a wave-fixed scattering potential is also considered, associated with the breaking of the scale-invariance symmetry, as a new additional mechanism for otoacoustic emission generation. A good fit was obtained, across stimulus level and frequency, for roughness, and not for nonlinear distortion, nor for scale-invariance violation. The phase-gradient delay of the same transient evoked otoacoustic emissions was consistent with the latency measured using a wavelet time-frequency technique, at all stimulus levels and frequencies. The results suggest that cochlear reflectivity is dominated by a component with a rapidly rotating phase, at all stimulus levels, in apparent contradiction with the usual assumption that, at high stimulus levels, a significant contribution to the transient evoked otoacoustic response should come from nonlinear distortion. Possible interpretations of this phenomenology are critically reviewed and discussed, considering the theoretical uncertainties and the limitations of the experimental technique. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2990711]

PACS number(s): 43.64.Jb, 43.64.Kc [BLM] Pages: 2995–3008

I. INTRODUCTION

Otoacoustic emissions (OAEs) are acoustic signals generated in the cochlea as a by-product of its active, nonlinear amplification mechanism (Probst *et al.*, 1991). Transient evoked OAEs (TEOAEs) are measured using a broadband, short-duration, transient stimulus, which can be a short pulse (click), a tone burst, or a maximum length sequence (Hine *et al.*, 2001).

Cochlear models based on the transmission-line formalism (e.g., Furst and Lapid, 1988; Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera *et al.*, 2005) represent the signal propagation along the basilar membrane (BM) as a traveling wave (TW). For each frequency component of the stimulus, forward propagation continues until the wave packet approaches its tonotopic place, where it produces the maximum transversal displacement of the BM (maximum of the activity pattern). In that cochlear region, a backward wave (OAE) at that frequency can be generated. No significant forward propagation occurs beyond the tonotopic place, due to local resonant absorption of that frequency component of the TW. According to the above-mentioned models, we will assume that OAEs are generated near the resonance place and propagate toward the base along the BM as a backward TW, neglecting other possibilities (for a discussion of different models for the backward propagation of OAEs, see, e.g., Shera *et al.*, 2007 and Moleti and Sisto, 2008).

Perturbative analyses based on the osculating parameter formalism applied to the WKB solution in a one-dimensional (1D) transmission-line model (Shera and Zweig, 1991) can be used to estimate the cochlear reflectivity associated with different mechanisms (Talmadge *et al.*, 2000). Indeed, several mechanisms should be considered for the OAE generation, and, among them, the more frequently discussed are nonlinear distortion and linear reflection from randomly distributed perturbations (Shera and Guinan, 1999).

When a given cochlear place is excited to a sufficiently high level that the BM displacement enters a nonlinear regime, the nonlinearity acts as a source (forcing) term in the cochlear differential equations, and produces emission of backward and forward TWs from that cochlear place. When a cochlear place is simultaneously excited by a complex stimulus, the system nonlinearity also generates frequencies that are linear combinations of those present in the stimulus. These phenomena are the primary mechanisms for distortion product OAE (DPOAE) generation and have been demonstrated to contribute also to the TEOAE response at high stimulus levels in guinea pigs (Withnell and McKinley, 2005) and humans (Withnell *et al.*, 2008). As nonlinear distortion occurs at cochlear places where the transverse BM displacement is large, the generation place is bound to be

[a)]Electronic mail: renata.sisto@ispesl.it
[b)]Electronic mail: moleti@roma2.infn.it

near the tonotopic place of the stimulus frequency or, as in the DPOAE case, near a place that is a univocal function of the frequencies of the stimulus. Therefore, nonlinear distortion generation is considered a "wave-fixed" mechanism.

A different OAE generation mechanism is associated with the linear reflection of a fraction of the TW in proximity of the tonotopic place. According to the coherent reflection filtering (CRF) theory, micromechanical irregularities of the cochlear structure, randomly distributed along the BM, which act as backscattering sources for the forward wave (Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera *et al.* 2005), are necessary to get significant reflection through a coherent mechanism. The perturbative analysis shows that the amplitude of each backscattered wave is significant only where the BM displacement is large. Therefore, the "center of mass" of the OAE source is in any case close to the tonotopic place for each stimulus frequency. On the other hand, the backscattering sources are localized at fixed positions that do not individually change with the stimulus frequency, so the mechanism is usually considered "place-fixed," as regards the frequency dependence of the overall reflectivity phase.

A different theoretical approach, recently proposed by Choi *et al.* (2008), which considers a 1D transmission-line model with impedance irregularities, suggests that a substantial fraction of the stimulus-frequency OAE (SFOAE) response could be due to linear reflection from a cochlear region remote from the peak of the TW, and much closer to the base.

In a perturbative approach, the scale-invariance violation associated with the experimentally observed variation of tuning with frequency (e.g., Sisto and Moleti, 2007) could also be responsible for wave-fixed generation of OAEs, as will be discussed later in this study. Another generation mechanism of the wave-fixed type had been proposed (Sisto and Moleti, 2005), in which the reflection of a fraction of the incoming wave is due instead to the sharp transversal impedance variation experienced by each frequency component approaching its own resonant place. Both analytical solutions of passive cochlear models (de Boer and McKay, 1980) and perturbative analyses based on the osculating parameter formalism applied to the WKB solution in 1D transmission-line models (Shera and Zweig, 1991) predict no appreciable reflections from such a mechanism, essentially because resonant absorption prevails in the same region where reflections should originate.

In a recent study, the characteristic times of the click-evoked OAE response have been compared, to get information about the place-fixed or wave-fixed nature of the TEOAE generation mechanisms (Sisto *et al.*, 2007). The click stimulus is well localized in the time domain, permitting one to define the latency in the time domain as the interval between the click time and the time of the maximum of the OAE response at a given frequency, measured with time-frequency analysis based on the wavelet transform (Tognola *et al.*, 1997). For TEOAEs, it is also possible to measure the phase-gradient delay by analyzing the Fourier transform (FT) of the pulse response. The direct comparison between the two different characteristic delays obtained by analyzing the same data has shown a close correspondence between them (Sisto *et al.*, 2007).

Withnell *et al.* (2008) recently demonstrated the presence of a short-latency, slow-rotating phase component in TEOAEs, which is typically canceled by windowing of the response waveform. This component is relatively more important at higher stimulus levels, consistently with the hypothesis of nonlinear distortion generation. Time-domain filtering was used to separate this slow-rotating component from the main contribution, whose phase-gradient delay was consistent, for each frequency, with cochlear round-trip delay. Notaro *et al.* (2007) applied a matching pursuit algorithm (Jedrzejczak, 2004) to TEOAE analysis to identify, within a peculiar class of "atoms" with short latency and low frequency, those whose frequency matched that expected for intermodulation distortion products of two higher frequency atoms of the same TEOAE. The matching pursuit analysis decomposes the waveform into a set of atoms, which are wave packets of definite central frequency, duration, and latency. It turned out that a substantial fraction of the short-latency, low-frequency atoms were attributable to intermodulation distortion. Due to the acquisition window used by Notaro *et al.* (2007), which cut off the first 2.5 ms of the TEOAE response, their observation could not be checked at higher frequency. These results suggest that the short-latency OAE contribution found by Withnell *et al.* (2008) probably comes from intermodulation distortion and not from nonlinear generation at the same frequency of the stimulus, at its tonotopic place. Intermodulation distortion products would have much shorter latency than TEOAEs of the same frequency generated at their tonotopic place (by any mechanism), because their source is more basal than their resonant place, and therefore their forward path is shorter and their backward path is both shorter and faster.

We remark here that time-frequency estimates of the TEOAE latency are a direct measure of the physical delay between the time of the transient stimulus and the OAE response within a given frequency band (Sisto *et al.*, 2007). The wavelet coefficient amplitude squared at a given time and frequency represents the energy density within the corresponding "tile" in the time-frequency domain. Therefore the time of the maximum of the wavelet coefficient of a given frequency is a direct measure of the transmission delay of the most intense energy pulse in that frequency range. For matching pursuit, the interpretation of the delay of each wave packet is even more straightforward.

The physical meaning of the characteristic times associated with the phase gradient is less obvious and should be discussed. Let us momentarily forget the fact that the TEOAE response is highly nonlinear, and that, particularly, using the nonlinear mode of acquisition, we actually cancel any linear component. It must be stressed that this does not mean at all that we are sensitive only to nonlinear distortion components. In fact, linearity of the generation mechanism does not imply linearity of the input/output (I/O) TEOAE function, which would result in a null "nonlinear mode" signal. Anyway, only for linear systems the full correspondence between the frequency and time domain representations is

guaranteed, and one can define a complex response function, which is the FT of the impulse response. If the TEOAE response to a click is considered as an experimental approximated evaluation of the pulse response of a linear system, then its FT would be an approximation of the system frequency response, and the derivative of its phase with respect to frequency would provide an estimate of the physical latency of the response (again, as a function of frequency). As a consequence, as wave-fixed OAEs are predicted to have a constant phase, they should have zero latency. A contradiction arises for TEOAEs, whose latency must obey causality requirements, associated with the finite delay needed for the transmission of the acoustic signal along the BM. In literature, the constant-phase component of the OAE response is often called short latency, meaning not only that its phase-gradient delay is null, but also assuming that the response is instantaneous. This is not possible, if one assumes that the emission place coincides with the tonotopic place, for each frequency, also considering that independent estimates of the forward transmission latency from derived auditory brainstem response (ABR) data set lower limits of a few milliseconds to the OAE latency (see, e.g., Moleti and Sisto, 2008). On the other hand, as mentioned above, the TEOAE contributions from intermodulation distortion products, which are generated, for each DP frequency, at a more basal place than the resonant place, could escape this requirement, because their forward delay would be shorter than expected for that frequency, and, more importantly, their backward delay would be much shorter, because the transmission delay of the backward wave would not include the slowest part of the path near the resonant place. In this rather uncertain theoretical framework, experimental studies are necessary to provide accurate latency data at different stimulus levels using different techniques and algorithms.

As previously mentioned, Choi *et al.* (2008) recently proposed a model in which both long and short phase-gradient delay SFOAE components are generated by cochlear irregularities, respectively, from the resonant region and from a much more basal region. In this case, the presence of a short-delay OAE component would not require assuming any wave-fixed generation mechanism, and some TEOAE experiments could fail to detect it, simply due to delayed onset of the acquisition window.

In this study, we present a new data analysis, in which both TEOAE I/O functions and latencies are considered. In fact, different OAE generation mechanisms predict different dependences of the TEOAE response on the stimulus level. The experimental TEOAE I/O curves have been compared with analytical predictions from different theoretical OAE generation models, using the osculating parameter formalism (Talmadge *et al.*, 1998; Talmadge *et al.*, 2000), and the results of this comparison are discussed. The comparison between phase-gradient delays and wavelet latencies, already presented in Sisto *et al.*, 2007, has been refined, also discussing the effect of arbitrary data analysis choices.

II. METHODS

TEOAEs of 14 young subjects (23–30 years), already analyzed in previous studies (Sisto and Moleti, 2007; Sisto *et al.*, 2007), recorded varying the click stimulus between 60 and 90 dB, using the ILO Echoport 292 (Otodynamics, Ltd.) system in the nonlinear acquisition mode, were analyzed in this study. A standard window (onset at 2.5 ms and flattop from 5 ms) was used to cancel the residual ringing artifact. This window also affects the high-frequency response above 4–5 kHz, as well as any other short-latency OAE response. The TEOAE complex FT of each ear has been divided by the FT of the stimulus, for each stimulus level, to get an estimate of the TEOAE complex transfer function (TF, or I/O function). As in a previous study (Sisto *et al.*, 2007) the phase-gradient delay was evaluated by differentiating the phase of this transfer function, which approximately coincides with that of the FT of the evoked waveform if the click stimulus is a good approximation to a delta function at zero delay. Due to limited acquisition time ($T=20$ ms), this operation yields a phase-gradient delay estimate mod(T), and this ambiguity was solved by forcing the values to belong to an arbitrarily chosen latency range, for comparison with the wavelet latencies. The latency of each OAE frequency component was evaluated using a time-frequency technique based on the wavelet transform (Sisto and Moleti, 2007).

III. MODEL

A. WKB approximation and 1D cochlear model

The osculating parameter method, applied to the WKB solution of 1D transmission-line model (Shera and Zweig, 1991), permits to evaluate perturbatively the contributions to cochlear reflectivity associated with different OAE generation mechanisms. The local difference between the complete equation and that for which the WKB approximation is the exact solution can be considered a source of reflected waves, distributed along the BM. The linearized equations in the frequency domain for a resonant 1D transmission-line model can be written as

$$\frac{\partial^2 P_d(x,\omega)}{\partial x^2} + k^2 P_d(x,\omega) = 0, \tag{1}$$

where x is the longitudinal position measured along the BM starting from the base, ω is the angular frequency of the Fourier component of the TW, k is the wave vector, function of ω and x, and P_d is the differential pressure applied to the BM.

The wave vector is tonotopically resonant as follows:

$$k^2 = \frac{k_0^2 \omega^2}{\Delta(x,\omega)}, \tag{2}$$

where $k_0 = 31\ \text{cm}^{-1}$ is a constant dependent on geometrical and structural cochlear properties, and Δ is a tonotopically resonant denominator (Talmadge *et al.*, 1998). It can include delayed stiffness terms (Zweig and Shera, 1995; Talmadge *et al.*, 1998), or, in the simplest approximation, be the simple resonant denominator associated with a passive resonance of effective width Γ as follows:

$$\Delta(x,\omega) = \omega_0^2(x) - \omega^2 + j\omega\Gamma(x). \tag{3}$$

The local resonance frequency is related to cochlear position by the Greenwood map (Greenwood, 1990)

$$\omega_0(x) = \omega_{\max} e^{-k_\omega x} + \omega_1, \tag{4}$$

where k_ω is the cochlear inverse scale length, and ω_1 is a small scale-invariance violating term, which is important only at low frequency (below 1 kHz) and will be neglected in the following. For the damping constant Γ we assume a similar exponential dependence on x, with an inverse length scale k_Γ.

In the WKB approximation, the solution is a superposition of basis functions that are progressive and regressive waves. The WKB basis functions, $\beta_\pm$, are the exact solutions of the equations

$$\frac{\partial^2 \beta_\pm}{\partial x^2} + k^2(1+\varepsilon_\pm)\beta_\pm(x,\omega) = 0. \tag{5}$$

Different basis functions can be used, yielding different expressions for the functions $\varepsilon_\pm$. Choosing functions of the type

$$\beta_\pm(x,\omega) \approx \frac{1}{\sqrt{k}} e^{\mp j\int_0^x k(x',\omega)dx'}, \tag{6}$$

which are those usually defined as the WKB basis functions (Shera and Zweig, 1991), we obtain

$$\varepsilon_\pm = -\frac{1}{k^2}\sqrt{k}\frac{\partial^2}{\partial x^2}\frac{1}{\sqrt{k}}. \tag{7}$$

If ε is sufficiently small, a perturbative regime holds, and the basis functions are also good approximated solutions of Eq. (1).

B. Derivation of cochlear reflectivity

The exact solution can be developed on the basis of the WKB basis functions

$$P_d = P_+ + P_- = \psi_+\beta_+ + \psi_-\beta_-, \tag{8}$$

where P_+ represents a progressive wave traveling toward the apex, whereas P_- represents a regressive wave traveling back toward the cochlear base. The coefficients of the basis functions (called osculating parameters) are functions that can be interpreted as amplitudes of the progressive and regressive waves. The amplitude at the cochlear position x for a regressive wave generated by the reflection of a progressive wave of unitary amplitude (reflection coefficient) is

$$\psi_- \cong \int_x^\infty dx' \beta_+\sigma_+\beta_+, \tag{9}$$

where σ_+ can be defined as a perturbative potential locally responsible for the reflection, which is given by

$$\sigma_+ = k^2\gamma\varepsilon_+. \tag{10}$$

The function γ is the inverse of the Wronskian determinant of the solution

$$\gamma = \frac{1}{2jk(0)}, \tag{11}$$

with

$$k(0) = \frac{k_0\omega}{\omega_{\max}}. \tag{12}$$

From Eq. (7) one can appreciate that the rapid change of the wave vector near the resonant place is, for each frequency component of the forward TW, a backscattering source, whose integral equation (9) is, however, negligibly small, as previously mentioned. Therefore, to explain the observed OAE levels, additional backscattering sources have been considered, among them the most important are nonlinear distortion and linear reflection from cochlear irregularities (e.g., Talmadge *et al.*, 2000). Addition of these sources may affect the validity of the frequency representation, which is, in any case, still valid at a sufficiently low stimulus level.

The basis function can be treated as a wave packet with wave vector near $\hat{k}$ and localized in the resonant place $\hat{x}$, with a width dependent on the Q of the resonance.

$$\beta_\pm(x,\omega) \approx \sqrt{\frac{k_0\omega}{\omega_{\max}}}\frac{1}{\sqrt{\hat{k}}} e^{-j\hat{k}(x-\hat{x})} e^{(x-\hat{x})^2/2\sigma_x^2} e^{-2j\pi N_\lambda} e^{-j\pi/2}. \tag{13}$$

Displacement basis functions are defined as

$$\chi_\pm(x,\omega) = \frac{\Delta(0,\omega)}{\Delta(x,\omega)}\beta_\pm = \frac{k^2(x,\omega)}{k^2(0,\omega)}\beta_\pm. \tag{14}$$

We considered a saturated regime in which the BM gain is given by

$$\hat{\chi}_+ = \eta(p/p_0)^a \hat{k}^{3/2}/k^{3/2}(0) = \eta(p/p_0)^a (k_0\sqrt{Q}/k(0))^{3/2}, \tag{15}$$

where p/p_0 is the stimulus amplitude in units of the reference pressure p_0=20 μPa, or 0.02 dyne/cm^2. The parameters η and a are introduced to parametrize, respectively, the size of the BM gain and its dependence on the stimulus level. The scaling with a power a of the stimulus amplitude keeps into account the fact that the cochlear amplifier gain is partly saturated at the stimulus levels considered in this experiment. In principle, the parameter a could vary in the range between 0 and $-2/3$ (corresponding, respectively, to linear amplification and to fully saturated behavior for a cubic nonlinearity). For η=35 and $a=-1/3$, we obtain, in the frequency and stimulus range of our experiment, reasonable values of the BM gain $\hat{\chi}$ between 40 and 60 dB, decreasing with stimulus level and, slowly, with frequency. This choice leads to a significant simplification of the theoretical expressions for the reflectivity. The dependence of Q on the stimulus level was derived from TEOAE latency in a previous study, in which the same data had been analyzed (Sisto and Moleti, 2007). These assumptions are not obvious and could be questioned. They mean that, for each frequency, the local quality factor affects both the TEOAE latency and the maximum BM transverse displacement through the local value of the factor $\hat{k}$ at the resonant place, while the overall dependence of the TEOAE response on the stimulus level is separately assumed to be governed by a general saturated behavior associated with the cochlear nonlinearity.

We expressed the activity pattern width σ_x in terms of the cochlear tuning Q as follows:

$$\sigma_x = \frac{\alpha}{k_\omega Q(f)}, \tag{16}$$

where $\alpha=1$ corresponds to a simple passive resonator, and used the approximate estimate of the wave vector amplitude at the resonant place

$$\hat{k} = k_0\sqrt{Q(f)}. \tag{17}$$

To compare the cochlear reflectivity estimates to the experimental TEOAE TF, one should take into account the middle ear I/O matrix. If we neglect it, we assume it is independent of frequency and stimulus level, and that in the forward and reverse conversions between pressure in the outer ear and differential pressure on the BM, the overall gain is 1. This is surely an excessive simplification, because it is rather well established that the round-trip middle ear gain is negative, and decreasing with increasing frequency in the 1–4 kHz range. This middle ear characteristic was invoked to explain the typical TEOAE spectral shape (Puria, 2003). Therefore one must keep into account this effect, and, when evaluating nonlinear generation mechanisms, one should separately consider the forward and reverse gains.

C. Nonlinear distortion reflectivity

The wave-fixed sources of reflection considered as plausible candidates in literature are associated with nonlinear distortion. In Appendix A, a complete calculation of the nonlinear distortion contribution to the spectral reflectivity in the perturbative limit (external P_{dr} close to the auditory threshold) is reported. These calculations, as most of those of the other appendixes, are not original but follow the scheme of Talmadge *et al.* (1998, 2000). We add them in Appendixes A, B, and C to help the interested reader through the derivation of the results. Assuming a nonlinear damping of the Van der Pol type

$$\Gamma = \gamma_0 + \gamma_2\xi^2, \tag{18}$$

the result [as in Eq. (29), p. 2916 of Talmadge *et al.*, 2000] is

$$\psi_{\mathrm{nl}} \cong -j\gamma\frac{\varepsilon_\gamma\omega^4 k_0^2|b_+|^2|\hat{\chi}_+|^4}{b_{\mathrm{nl}}^2\omega_{\max}^4}e^{-4j\pi N_\lambda}e^{-j\pi}e^{j\hat{k}k_\omega\sigma_x^2/2}\sigma_x\sqrt{\frac{\pi}{2}} \times\exp\left[-\left(\hat{k}^2-\frac{k_\omega^2}{4}\right)\frac{\sigma_x^2}{2}\right]. \tag{19}$$

Here, b_{nl}^2 is a cochlear displacement saturation level, ε_γ is the inverse of tuning Q, assumed to be constant, σ_x is the width of the cochlear activity pattern, N_λ is the number of wavelengths from the base to the tonotopic place, b_+ is the cochlear displacement at the base, and a hat above a function indicates that it is evaluated at the resonant place.

The nonlinear distortion reflectivity function amplitude is

$$|\psi_{\mathrm{nl}}| \cong \sqrt{\frac{\pi}{2}}\frac{k_0}{k_\omega}\frac{a_+^2}{2b_{\mathrm{nl}}^2\sigma_{\mathrm{BM}}^2\omega_{\max}}\eta^4\left(\frac{p}{p_0}\right)^{2a}\alpha\frac{Q}{\omega^3} \times\exp\left[-\frac{\alpha^2k_0^2}{2Qk_\omega^2}\right], \tag{20}$$

where a_+ is the differential pressure amplitude at the base, assumed proportional to the stimulus amplitude. Note that this expression contains an explicit dependence on frequency in the factor Q^2/ω^3, but its stronger dependence on frequency is due to the factor Q in the argument of the exponential function, if its absolute value is larger than unity, and if Q increases with frequency, which is typically true.

D. Roughness reflectivity

The reflection sources responsible for the rapid-rotating phase component of the OAE response are generally presumed to be associated with small irregularities randomly distributed along the BM. This cochlear "roughness" would contribute to the cochlear reflectivity through the so-called CRF mechanism (Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera *et al.*, 2005).

The computation of the reflectivity contribution due to roughness is shown in detail in Appendix B. The calculation reproduces the results by Talmadge *et al.* (1998).

$$\psi_r = -\frac{\omega^4k_0^2\hat{\chi}_+^2 2r_0}{2jk(0)\omega_0^4}e^{-4j\pi N_\lambda}e^{-j\pi}\sigma_x e^{k_\omega^2\sigma_x^2}e^{2j\hat{k}\hat{x}}\sqrt{\pi}\tilde{r}_f(\hat{x}-k_\omega\sigma_x^2). \tag{21}$$

As in the case of the nonlinear distortion mechanism, the original contribution of this study is essentially having made explicit the dependence of Q on the results, and the use of Eq. (15). The amplitude of the reflectivity function is

$$|\psi_r| = \frac{\omega^4k_0^2|\hat{\chi}_+|^2 2r_0}{2k(0)\omega_0^4}\sigma_x e^{k_\omega^2\sigma_x^2}\sqrt{\pi}|\tilde{r}_f(\hat{x}-k_\omega\sigma_x^2)| = \sqrt{\pi}r_0|\tilde{r}_f|\eta^2\alpha\frac{k_0}{k_\omega}\left(\frac{p}{p_0}\right)^{2a}Q^{1/2}e^{\alpha^2/Q^2}, \tag{22}$$

where $r_0|\tilde{r}_f|$ parametrizes the cochlear roughness amplitude, given by stiffness fluctuations of the form

$$\tilde{\omega}_0(x) = \omega_0(x)(1+r(x)\tilde{r}(x)). \tag{23}$$

It is interesting to note that having assumed the validity of Eq. (15) leads to cancellation of the explicit dependence of frequency on the reflectivity estimate for roughness. The only frequency dependence is through that of cochlear tuning Q.

E. Scale-invariance violation reflectivity

A new reflection source was also considered, namely, a linear term coming from a perturbative scale-invariance violation associated with the frequency dependence of cochlear tuning. A detailed calculation is presented in Appendix C. The result is

$$\psi_{\text{siv}} \cong j\gamma \frac{d\varepsilon_\gamma \omega^4 k_0^2 |\hat{\chi}_+|^2}{\Delta(0)^2} e^{-4j\pi N_\lambda} e^{-j\pi} e^{j\hat{k}k_\omega \sigma_x^2} \times \sigma_x \left(\hat{x} - \sigma_x^2\left(j\hat{k} + \frac{k_\omega}{2}\right)\right)\sqrt{\pi}\exp\left[-\left(\hat{k}^2 - \frac{k_\omega^2}{4}\right)\sigma_x^2\right], \tag{24}$$

where the scaling symmetry breaking is parametrized by the value of $d=k_\omega - k_\Gamma$, assuming for the damping coefficient a Greenwood-like exponential dependence on x, with a different inverse length scale k_Γ. The phase is the same as that in Eq. (19) and it is a slowly varying function of frequency. The amplitude of the reflectivity is, approximately,

$$|\psi_{\text{siv}}| \cong \frac{\sqrt{\pi}k_0}{2k_\omega} d\varepsilon_\gamma \hat{x}\eta^2 \left(\frac{p}{p_0}\right)^{2a} \alpha Q^{1/2} \exp\left[-\frac{k_0^2\alpha^2}{k_\omega^2 Q}\right]. \tag{25}$$

In this case, as for Eq. (22), the explicit dependence on frequency disappears, apart from a slow (logarithmic) dependence on the factor $\hat{x}$. In both cases, the frequency dependence of reflectivity is mainly due to the frequency dependence of the quality factor Q.

F. Phase-gradient delay and cochlear delay for different OAE generation mechanisms

For a 1D tonotopically resonant transmission-line cochlear model, the latency of a particular Fourier component of the response to an impulsive stimulus, which we measure with wavelet analysis, is given by

$$\tau_W(\omega) = 2\int_0^{\hat{x}} \frac{\partial\, \text{Re}(k)}{\partial\omega} dx. \tag{26}$$

The only assumption is done that the wavelet latency is the physical delay associated with the propagation of a wave packet of central frequency ω, which propagates along the BM with its group velocity.

The OAE response phase, in the WKB approximation, is

$$\phi(\omega) = -2\int_0^{\hat{x}} \text{Re}(k) dx, \tag{27}$$

having considered (as in Sisto *et al.*, 2007) that in the case of OAE measurement we are interested in the phase of the pressure in $x=0$, where the phase of the multiplicative external factor of the WKB functions, $k^{-1/2}$, is close to zero.

On this basis, Sisto *et al.* (2007) derived approximate expressions for the expected relation between phase-gradient delay and wavelet latency. For place-fixed OAE generation mechanisms, the experimentally measured phase-gradient delay would be

$$\tau_{\text{pg}}(\omega) = -\frac{d\phi}{d\omega} = 2\int_0^{\hat{x}} \frac{\partial\, \text{Re}(k)}{\partial\omega} dx = \tau_W(\omega), \tag{28}$$

whereas, for wave-fixed generation mechanisms, the approximate relation would be valid as follows:

$$\tau_{\text{pg}}(\omega) = -\frac{\partial\phi}{\partial\omega} - \frac{\partial\phi}{\partial\hat{x}}\frac{\partial\hat{x}}{\partial\omega} = 2\int_0^{\hat{x}} \frac{\partial\, \text{Re}(k)}{\partial\omega} dx + 2\,\text{Re}(\hat{k})\frac{\partial\hat{x}}{\partial\omega} \approx \tau_W(\omega) - \beta\frac{k_0\sqrt{Q}}{\omega k_\omega}, \tag{29}$$

where β is an dimensionless model-dependent constant of order 1, and where it has been considered that the phase-gradient delay is actually calculated as a total derivative of the phase with respect to the frequency, so, in the wave-fixed case, the derivative of the integration upper limit must also be considered.

In scale-invariant models, the wave-fixed prediction is that the phase-gradient delay is close to zero. If the scale-invariance symmetry is broken, the phase-gradient delay can be either positive or negative, but, in any case, should be much smaller than that predicted by place-fixed generation models.

If we assume that the OAE backward propagation is due to slow transverse TWs on the BM, and define the latency as the physical time that the signal needs to reach the tonotopic place and to come back to the base of the cochlea, for causality reasons, a short-latency component of the TEOAE signal simply could not exist. As mentioned in the Sec. I, independent estimates of the forward cochlear delay from derived ABR measurements imply forward delays of several milliseconds at 1–2 kHz. The signal at a particular frequency must travel from the base to the peak region and back from the activity pattern region to the base before it can be detected in the ear canal. On the other hand, a hypothetical wave-fixed, scale-invariant, generation mechanism, in an approximately linear cochlea, should produce OAEs with null phase-gradient delay, i.e., instantaneous response, if the system nonlinearity is sufficiently small, which seems to imply a little paradox. This issue deserves to be clarified. Thus, we will proceed presenting the results, leaving their interpretation open to further theoretical contributions.

IV. RESULTS

A. TEOAE transfer function, middle ear round-trip gain, and cochlear reflectivity

The amplitude of the experimental TEOAE transfer function is plotted in Fig. 1, for two ears of our sample, and in Fig. 2, where an average is performed over the 28 ears, to get a better view of the overall dependence on frequency and stimulus level. The results are similar to those shown by Kalluri and Shera (2007), and show a clear dependence on the stimulus level. The decrease in reflectivity with increasing stimulus level is an expected consequence of saturation of the cochlear amplifier. Individual data show quasiperiodic oscillations, similarly to what is observed in the SFOAE spectra (Kalluri and Shera, 2007), whereas the average data permit to appreciate the dependence on the stimulus level and frequency with higher precision.

The experimental TEOAE TF can be compared to theoretical estimates for the cochlear reflectivity only taking into account the forward and reverse transmission properties of the middle ear. Accurate measurements of the I/O character-

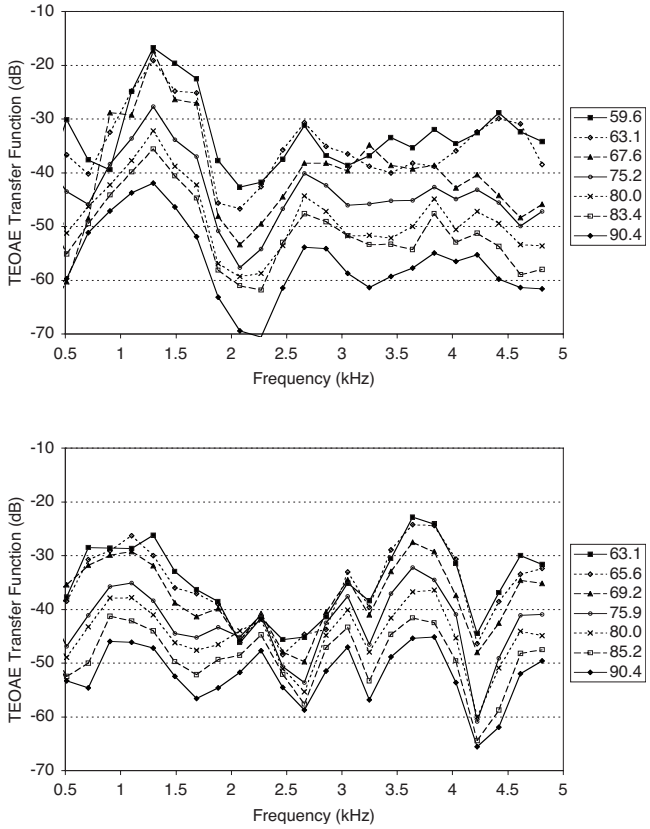


FIG. 1. Individual TEOAE TF function (amplitude) for two ears of the sample, for seven different stimulus levels (in dB pSPL). The dependence on the stimulus level is clearly visible, as well as significant individual spectral features and oscillations, similarly to what is observed in SFOAE TFs (Kalluri and Shera, 2007).

istics of the human middle ear have been obtained on cadavers by Puria (2003). Assuming that their results fully apply to living ears too, the middle ear forward gain (positive) and reverse gain (negative) must be taken into account separately in the interpretation of the TEOAE TF. Indeed, they are independent of the stimulus level, but frequency dependent, so the spectral shape of the TEOAE TF reflects also the middle ear spectral characteristics. The effect of the TEOAE TF on

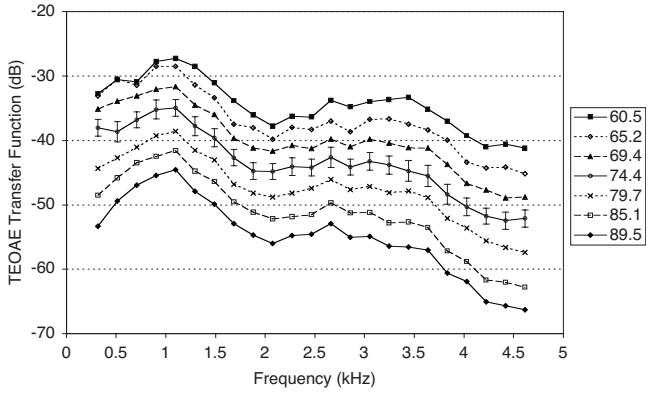


FIG. 2. TEOAE TF amplitude averaged over the 28 ears of the sample, for seven different stimulus levels. A monotonic dependence on the stimulus level is more clearly visible than in Fig. 1, as well as an overall decreasing trend with increasing frequency, which was obscured by individual spectral features in most ears. The standard error of the average is plotted for the middle stimulus level (74.4 dB pSPL, on average).

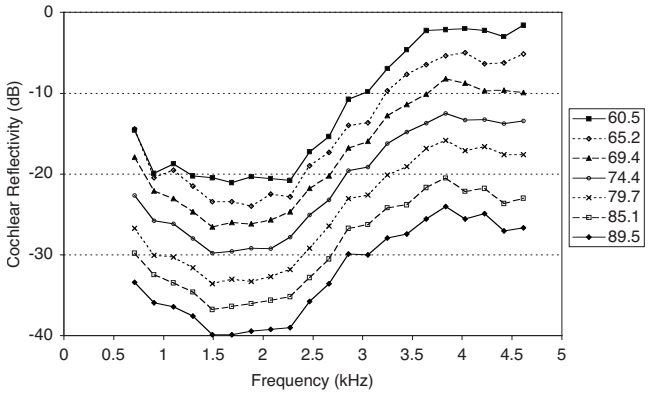


FIG. 3. Cochlear reflectivity evaluated from the TEOAE TF amplitude of Fig. 2, for seven different stimulus levels. The peculiar spectral behavior can be partly biased by having inaccurately estimated the frequency dependence of the middle ear round-trip gain.

the middle ear forward and reverse transmissions is synthesized by the round-trip middle ear pressure gain $G_{\mathrm{me}}^{\mathrm{RT}}$, which, according to Fig. 5 of Puria (2003), has a maximum of about −7 dB around 1.2 kHz, and decreases to −15 dB at 2 kHz and to −35 dB at 4 kHz. Cochlear reflectivity, estimated as

$$R(f,S) = \frac{\mathrm{TF}(f,S)}{G_{\mathrm{me}}^{\mathrm{RT}}(f)}, \tag{30}$$

is shown in Fig. 3, as a function of frequency at different stimulus levels. The experimental data show a peculiar spectral pattern, with a reflectivity minimum at intermediate frequencies. This result must be evaluated with some attention, because it may depend on an overestimate of the frequency dependence of $G_{\mathrm{me}}^{\mathrm{RT}}$, and the middle ear affects it in a subtler way. In fact, the frequency dependence of the middle ear forward transmission means that, using a click stimulus, which is spectrally flat in the ear canal, we are actually stimulating the cochlear base with a stimulus whose spectrum is peaked around 1 kHz, and decreases with frequency by approximately 6–7 dB/octave at higher and lower frequencies. Due to the dependence on the stimulus level shown in Fig. 3 itself, if we would account for this effect, and evaluate a reflectivity spectrum at constant stimulus spectral density at the cochlea by interpolation of the seven curves of Fig. 3, we would get a much weaker frequency dependence.

The dependence of cochlear reflectivity on frequency and stimulus level shown in Fig. 3 has been fitted to different theoretical expressions relative to different OAE generation mechanisms, in four selected 200-Hz-wide frequency bands, centered around 1, 2, 3, and 4 kHz. In this comparison, the above-mentioned effect of the stimulus level on the forward middle ear gain has been accounted for in the computation of the theoretical reflectivity, by multiplying the stimulus pressure density by a function of frequency proportional to the middle ear forward gain in Eqs. (15), (20), (22), and (25).

B. Reflectivity from roughness, nonlinear distortion, and scaling symmetry violation

The dependence of cochlear reflectivity on stimulus amplitude, shown in Fig. 4, is roughly consistent with a power law, with an exponent between −0.55 and −0.7, slowly in-

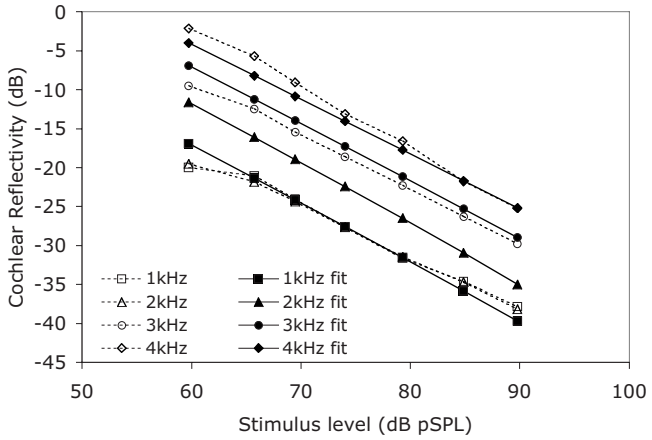


FIG. 4. Theoretical prediction for cochlear reflectivity due to roughness, Eq. (18) (white symbols), fitted to the 1, 2, 3, and 4 kHz experimental TEOAE TF data (black symbols). $r_0\tilde{r}_f=0.04\%$.

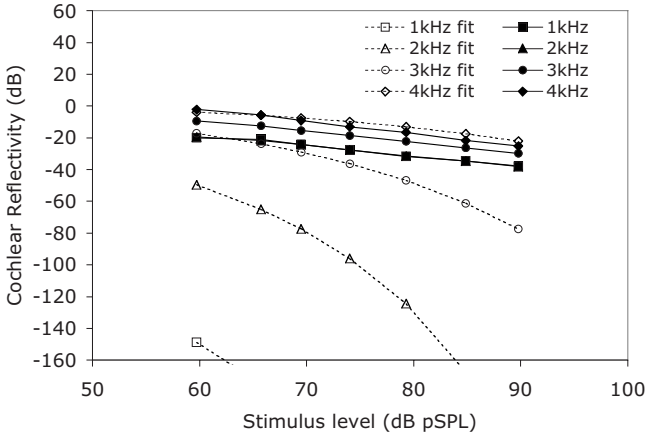


FIG. 5. Theoretical prediction for cochlear reflectivity from nonlinear distortion, Eq. (16) (white symbols), fitted to the 1, 2, 3, and 4 kHz experimental TEOAE TF data (black symbols). The fit is obtained at a single frequency, by increasing by 20 dB the cochlear gain with respect to that assumed in Fig. 4, but the dependence on frequency is completely in disagreement with the data. The dependence on both stimulus level and frequency could be approximately fitted only assuming unreasonably narrow cochlear activity patterns ($\alpha<0.1$).

creasing with frequency. This slope is rather steeper than what can be inferred from the slope of the SFOAE I/O curves reported by Choi *et al.* (2008), in an equivalent stimulus level range. A good fit to the theoretical prediction for roughness reflectivity, as regards the dependence on the stimulus level, is obtained for all frequencies, with the choice $a=-1/3$. Acceptable fit can also be obtained for different values of a, in the range between −0.4 and −0.25, and adjusting the parameter η to keep the BM gain in the same range. The best fit procedure with the theoretical expression for cochlear reflectivity due to roughness, Eq. (22), yields, for $r_0\tilde{r}_f=0.04\%$, the result shown in Fig. 4, for the frequency bands centered at 1, 2, 3, and 4 kHz. The dependence of reflectivity on tuning Q is rather slow. Due to the dependence on frequency and stimulus level of the experimental Q, the width of the activity pattern σ_x is also variable, in the range 0.03–0.17 cm. These values are somewhat smaller than those used by Talmadge *et al.* (2000). In a more refined model, as the inclusion of time-delayed stiffness terms effectively increases the width of the activity pattern, one could assume that σ_x is given by Eq. (16) with $\alpha>1$. In fact, the estimate of tuning used in this work is obtained from the TEOAE latency, and is therefore sensitive to the slowing down of the TW near the resonant place, which is more dependent on the steepness of the relation between wave vector and position than on the width of the resonance. The two properties are linked to the single parameter Q only for simple passive resonators. We remark that an acceptable fit has been obtained over the whole frequency range, for a single value of the roughness, letting σ_x vary with frequency only through the frequency dependence of the experimental cochlear tuning estimates, independently obtained from the same TEOAE data (Sisto and Moleti, 2007). A significant discrepancy is noted between 1 and 2 kHz, but one should also consider that the frequency dependence of the results is dramatically influenced by that of the functions assumed to describe the middle ear gain, which is not well known.

The nonlinear distortion mechanism predicts a peculiar dependence on the stimulus level and frequency, which cannot be fitted to the data. If one assumes $\alpha\geq 1$, a negligible reflectivity is predicted, due to the very small value of the exponential function in Eq. (20). Due to the strong dependence on the BM gain of Eq. (19), significant reflectivity could be obtained by increasing the parameter η, obtaining, for example, BM gains in the range 60–80 dB, as done in Fig. 5. In this case, however, it is not possible to get a good fit at all frequencies. A rough fit at all frequencies and stimulus levels could be obtained only assuming $\alpha<0.1$, which corresponds to very small values of the activity pattern $\sigma_x<0.01$ cm. Such small values of σ_x would not be coherent with the experimental cochlear tuning, and could not be easily reconciled with the accurate measurements of the BM transfer function that were the main motivation for Zweig (1991) to introduce the so-called time-delayed stiffness terms in the cochlear equations, obtaining, as a result, tall and broad activity patterns.

The fit to the theoretical expression for reflectivity due to scaling symmetry violation, Eq. (25), could be performed varying the parameter d associated with the degree of scaling symmetry breaking. No appreciable reflection is predicted if one assumes $\alpha\geq 1$, because, similarly to what happens for nonlinear distortion, the exponential factor in Eq. (25) is always negligibly small. A reasonably good fit could be obtained, again, only assuming unreasonably narrow activity patterns, as shown in Fig. 6, where the fit is obtained for $d=0.15$ and $\alpha=0.1$. For both wave-fixed mechanisms, the dependence on Q becomes unreasonably strong as the argument of the exponential is made of order unity, either by increasing Q or by decreasing α.

The two proposed wave-fixed mechanisms seem therefore unable to explain the observed global behavior of TEOAE I/O functions, without assuming a significant (dominant) contribution from roughness. A fit with a composite source could be possible, but too many parameters would be involved to get additional information.

C. Wavelet latency and phase-gradient delay

The comparison between the wavelet latency and the phase-gradient delay is not straightforwardly obtained, be-

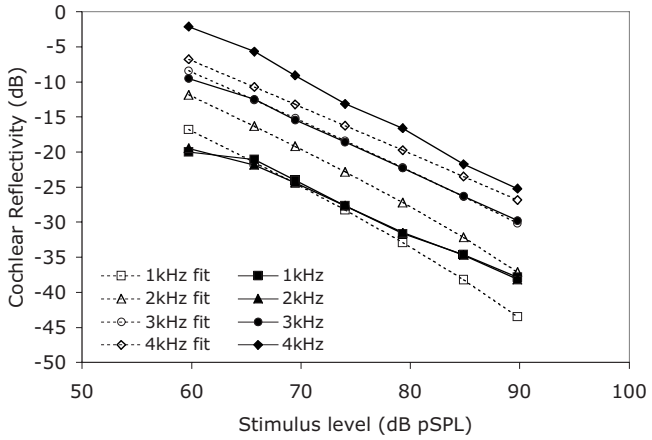


FIG. 6. Predicted cochlear reflectivity from scale-invariance violation, Eq. (20) (white symbols), fitted to the 1, 2, 3, and 4 kHz experimental TEOAE TF data (black symbols). The dependence on frequency and stimulus level may be approximately fitted, for low values of the scaling symmetry breaking parameter, only assuming unreasonably narrow cochlear activity patterns (α=0.1 and d=0.15 in the fit).

cause the phase-gradient delay is measured mod(T), where T is the click repetition rate (20 ms in our measurements). The removal of this ambiguity requires arbitrary assumptions (Sisto *et al.*, 2007). The apparently simplest assumption would be to transpose the experimental values in the interval $0-T$. The result of this procedure is shown in the top panel of Fig. 7, where the wavelet latency values (solid lines) are compared to the phase-gradient delays obtained in this way. The agreement is rather good below 2 kHz, but it is evident that at higher frequencies the phase-gradient delay becomes significantly longer than the wavelet latency, particularly at the lowest stimulus levels. If we force τ_{pg} to be in a range of width $T/2$ around $\tau(f)$, with $\tau(f)=af^b$, we obtain the result shown in the mid- and bottom panels of Fig. 7, using $a=T/2$ and, respectively, $b=-0.5$ and -1. The choice $b=-0.5$ roughly corresponds to a fit to the wavelet latency measurements, so there is no surprise that doing so we get the best agreement between the two delays. On the other hand, it must be stressed that the apparently conservative choice $b=0$ is equally arbitrary, and may be responsible for a systematic error leading to the discrepancy shown in Fig. 6(a). In fact, the highest signal-to-noise ratio (SNR) is, for these data, in the range 1–2.5 kHz. Outside this range, for the spectral points with a low SNR, random phase values are obtained, which lead to random phase-gradient delay estimates, both positive and negative. If we force them to be in the range $0-T$, we obtain an average value tending to $T/2$ as SNR tends to zero, and a significant systematic overestimate of τ_{pg} in all frequency bands in which the "true" τ_{pg} is significantly lower than $T/2$, and an underestimate where the opposite happens. The rather flat phase-gradient delay curves shown in the top panel of Fig. 7 could be attributed to such a systematic effect. The choice $b=-0.5$ does not introduce any bias as long as the average τ_{pg} is in agreement with the fit to the wavelet latencies, whereas it does systematically drag τ_{pg} toward a fictitious agreement if it is not. It could be noted that the power law fit line lies systematically above the τ_{pg} data in the top panel and below them in the bottom one,

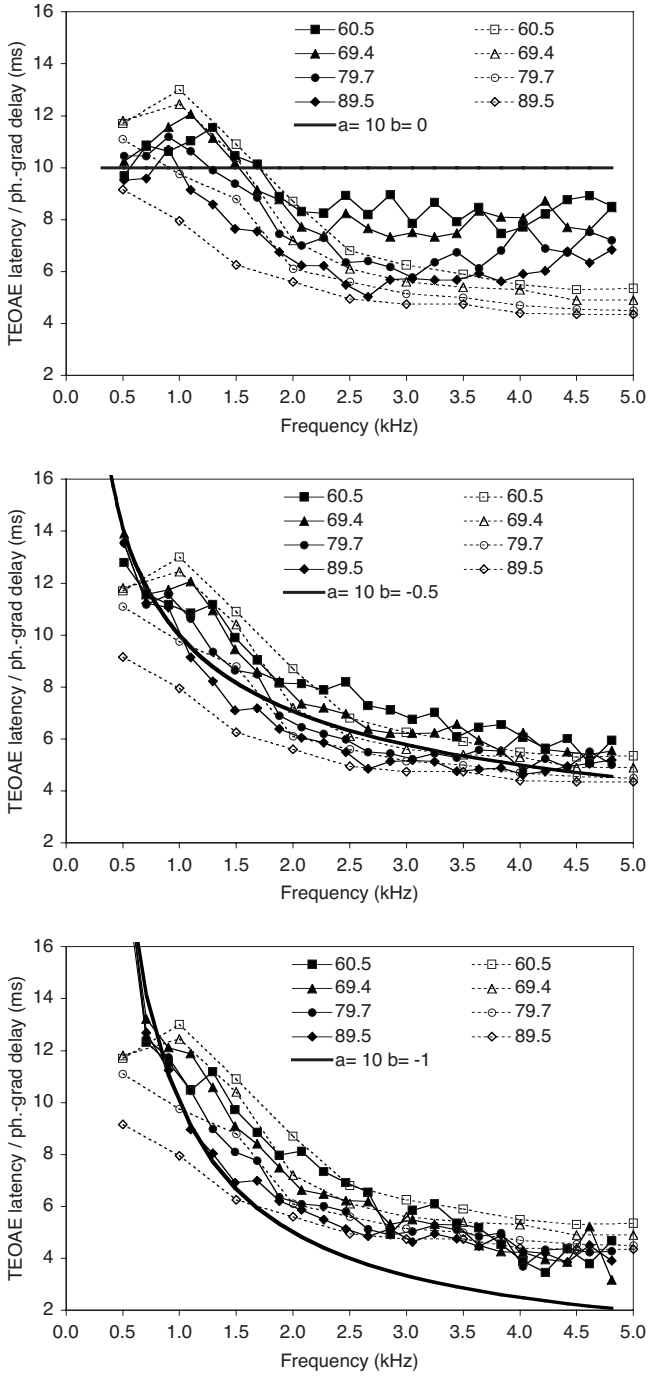


FIG. 7. Comparison between wavelet latency (dotted lines) and phase-gradient delay (solid lines). Phase-gradient delay is forced in the range $\tau(f)\pm T/2$, with $\tau(f)=af^b$, $a=T/2$, and $b=0$ (top panel). Same with $b=-0.5$ (mid) and $b=-1$ (bottom). The solid thick line represents the center of the accepted latency interval. The effect of the arbitrary choice of this interval may be appreciated. The choice of the midpanel approximately corresponds to centering the interval around the best fit to the wavelet latency data at intermediate stimulus levels.

suggesting that the "right" value is close to that used in the midpanel. In fact, hypothesizing a wrong delay interval drags the τ_{pg} estimates in that direction, but cannot lead to a good fit, if the data have a different intrinsic frequency dependence, whereas a good fit would be obtained for pure noise data for any choice of b. With all the caution required by the arbitrariness of the method, it can be concluded that the hy-

pothesis that wavelet latencies and phase-gradient delays are comparable, and exhibit a similar dependence on the stimulus level, is fully compatible with these results.

V. DISCUSSION

The OAE generation mechanisms are quite complex and their relative importance is dependent on the stimulus level. Linear reflection mechanisms associated with cochlear roughness are generally assumed to be responsible for the SFOAE and TEOAE responses at low stimulus levels, whereas there is some evidence for intermodulation distortion components in the TEOAE response, both in small mammals and in humans (Withnell and McKinley, 2005; Withnell *et al.*, 2008). This component, which cannot be evoked in the SFOAE response, could become dominant at a high stimulus level. In this case one should observe marked differentiation between the SFOAE and TEOAE behaviors as the stimulus level increases. A recent study (Kalluri and Shera, 2007) has demonstrated close correspondence, in a wide stimulus level range, 40 to 80 dB peak-equivalent SPL (pSPL), between SFOAE and TEOAE responses, if the stimulus levels are correctly "intercalibrated," suggesting that the same mechanisms should be responsible for both. Thus, nonlinear distortion probably plays a minor role in the generation of human TEOAEs.

This study aims at attacking the problem from a different point of view, comparing the amplitude and the phase of the TEOAE I/O functions with the predictions of theoretical models assuming different OAE generation mechanisms. The roughness reflectivity function fits well the amplitude data at all stimulus levels, using the width of the activity pattern σ_x derived from cochlear tuning estimates. A good fit cannot be obtained for the considered wave-fixed mechanism unless unreasonably narrow activity patterns are assumed.

The simplest interpretation of these results suggests that roughness could be the dominant source for TEOAE generation in the whole stimulus level range of this study. This was also recently suggested by Sisto *et al.* (2007), who observed that TEOAE phase-gradient delays and wavelet latencies tend to coincide at all stimulus levels between 60 and 90 dB pSPL. According to transmission-line cochlear models, the coincidence between the two delays would support indeed place-fixed OAE generation mechanisms. On the other hand, this last result may be partly biased by the TEOAE acquisition window. One possible explanation of the coincidence of the latencies is that, in the TEOAE case, the waveform window cancels any very early response, as that associated with intermodulation distortion between high-frequency primary components. It has also been suggested that, as the transient response is evoked by clicks of very short duration with respect to the acquisition time, a wave-fixed component could have a rapid-rotating phase (Talmadge *et al.*, 2000).

Summarizing, the results suggest that a wave-fixed source cannot be the only TEOAE source, but they cannot exclude some wave-fixed contribution at high stimulus levels, possibly hidden by the late acquisition window onset.

VI. CONCLUSIONS

This study confirms, by analyzing both the amplitude and the phase of the TEOAE I/O functions, that most of the TEOAE phenomenology can be explained by a place-fixed generation mechanism, associated with cochlear roughness, as predicted by the CRF theory (Shera *et al.*, 2005). Due to theoretical uncertainties, and to intrinsic limitations of the TEOAE acquisition technique used in this study (particularly, the window that cancels the earliest part of the TEOAE response), contributions from wave-fixed nonlinear generation sources cannot be fully excluded.

APPENDIX A: COCHLEAR REFLECTIVITY FROM NONLINEAR DISTORTION

Following Talmadge *et al.* (2000), we perturb the standard functional form of the wave vector

$$k^2(\omega,x) = \frac{k_0^2\omega^2}{\Delta(\omega,x)}. \tag{A1}$$

Assuming Van der Pol nonlinear damping, in the frequency domain we get

$$\Gamma(\omega,x) = \gamma_0(x) + \gamma_2(x)|\xi|^2, \tag{A2}$$

where the absolute value is due to the frequency domain approximation in which only the component of frequency ω is considered (Talmadge *et al.*, 2000), and where

$$\gamma_2(x) = \frac{\gamma_0(x)}{b_{\mathrm{nl}}^2} = \frac{\gamma_0}{b_{\mathrm{nl}}^2} e^{-k_\omega x} \equiv \frac{\varepsilon_\gamma}{b_{\mathrm{nl}}^2}\omega_0(x). \tag{A3}$$

Here it has been assumed that, as a first order approximation, damping has the same tonotopic dependence on x as frequency, i.e., that tuning is frequency independent, $Q = \varepsilon_\gamma^{-1}$. The resonant denominator takes the perturbed form

$$\tilde{\Delta}(x) = \omega^2(x) - \omega^2 + j\Gamma\omega = \omega^2(x) - \omega^2 + j\gamma_0\omega + j\omega\gamma_2\xi^2 = \Delta\left(1 + \frac{j\omega\gamma_2|\xi|^2}{\Delta}\right), \tag{A4}$$

$$\tilde{k}^2 = \frac{k_0^2\omega^2}{\tilde{\Delta}} = \frac{k_0^2\omega^2}{\Delta\left(1 + \frac{j\omega\gamma_2|\xi|^2}{\Delta}\right)} \cong \frac{k_0^2\omega^2}{\Delta}\left(1 - \frac{j\omega\gamma_2|\xi|^2}{\Delta}\right) = k^2\left(1 - \frac{j\omega\gamma_2|\xi|^2}{\Delta}\right). \tag{A5}$$

The perturbed wave vector differs from the unperturbed one by

$$\delta k_{\mathrm{nl}}^2 = k^2\frac{(-j\omega\gamma_2|\xi|^2)}{\Delta} = \frac{-j\gamma_2 k_0^2\omega^3|\xi|^2}{\Delta^2}. \tag{A6}$$

The reflectivity is given by the equation

$$\psi_{\mathrm{nl}} \cong \int_x^\infty dx'\beta_+\sigma_{\mathrm{nl}}\beta_+ \tag{A7}$$

with

$$\sigma_{\mathrm{nl}} = \delta k_{\mathrm{nl}}^2 \frac{1}{W(\omega)}. \tag{A8}$$

Treating the basis function as a wave packet with a wave vector near $\hat{k}$ and localized in $\hat{x}$, and using the displacement basis functions we get

$$\psi_{\mathrm{nl}} \cong -j\gamma \frac{\omega}{\omega^2 k_0^2} \int_x^\infty dx k^2(0) \chi_+ \gamma_2(x') |\xi|^2 k^2(0) \chi_+, \tag{A9}$$

$$\psi_{\mathrm{nl}} = -j\gamma \frac{\varepsilon_\gamma \omega^3 k_0^2}{b_{\mathrm{nl}}^2 \Delta(0)^2} \int_x^\infty dx' \chi_+^2(x',\omega) \omega_{\max} e^{-k_\omega x'} |\xi(x',\omega)|^2. \tag{A10}$$

On the other hand

$$k^2 P_d = k_0^2 \sigma_{\mathrm{BM}} \omega^2 \xi$$

$$\begin{aligned} \xi(x,\omega) &= \frac{k^2}{\sigma_{\mathrm{BM}} k_0^2 \omega^2} a_+ \beta_+ = \frac{k^2}{\sigma_{\mathrm{BM}} k_0^2 \omega^2} \frac{k^2(0)}{k^2} a_+ \chi_+ \\ &= \frac{k^2(0)}{\sigma_{\mathrm{BM}} k_0^2 \omega^2} a_+ \chi_+ = \frac{k_0^2 \omega^2}{\sigma_{\mathrm{BM}} \omega^2 k_0^2 \omega_{\max}^2} a_+ \chi_+ \\ &= \frac{1}{\sigma_{\mathrm{BM}} \omega_{\max}^2} a_+ \chi_+ = b_+ \chi_+, \end{aligned} \tag{A11}$$

where the relation between transverse displacement ξ and P_d is from Eq. (1) of Talmadge *et al.* (1998), σ_{BM} is the BM surface density in g cm^{-2}, and

$$b_+ = \frac{1}{\sigma_{\mathrm{BM}} \omega_{\max}^2} a_+. \tag{A12}$$

Substituting into Eq. (A10), one gets

$$\psi_{\mathrm{nl}} = -j\gamma \frac{\varepsilon_\gamma \omega^3 k_0^2}{b_{\mathrm{nl}}^2 \Delta(0)^2} \int_x^\infty dx' \chi_+^2(x',\omega) \omega_{\max} e^{-k_\omega x'} |b_+ \chi_+(x',\omega)|^2. \tag{A13}$$

Using a simplified, scale-invariant Greenwood map

$$\begin{aligned} \omega_0(x) &= \omega_{\max} e^{-k_\omega x} = \omega_{\max} e^{-k_\omega (x-\hat{x})} e^{-k_\omega \hat{x}} = \omega(\hat{x}) e^{-k_\omega (x-\hat{x})} \\ &= \omega e^{-k_\omega (x-\hat{x})}, \end{aligned} \tag{A14}$$

where $\omega(\hat{x}) = \omega$, by definition, and substituting in the integral, we get

$$\begin{aligned} \psi_{\mathrm{nl}} = &-j\gamma \frac{\varepsilon_\gamma \omega^4 k_0^2 |b_+|^2}{b_{\mathrm{nl}}^2 \omega_{\max}^4} \int_x^\infty dx' \chi_+^2(x',\omega) e^{-k_\omega (x'-\hat{x})} \\ &\times |\chi_+(x',\omega)|^2. \end{aligned} \tag{A15}$$

Developing the basis function

$$\begin{aligned} \chi_+(x,\omega) &\approx \sqrt{\frac{k(0)}{\hat{k}}} \frac{\hat{k}^2}{k^2(0)} e^{-j\hat{k}(x-\hat{x})} e^{(x-\hat{x})^2/2\sigma_x^2} e^{-2j\pi N_\lambda} e^{-j\pi/2} \\ &= \hat{\chi}_+ e^{-j\hat{k}(x-\hat{x})} e^{(x-\hat{x})^2/2\sigma_x^2} e^{-2j\pi N_\lambda} e^{-j\pi/2} \end{aligned} \tag{A16}$$

and substituting into Eq. (A15), we get

$$\begin{aligned} \psi_{\mathrm{nl}} = &-j\gamma \frac{\varepsilon_\gamma \omega^4 k_0^2 |b_+|^2 |\hat{\chi}_+|^4}{b_{\mathrm{nl}}^2 \omega_{\max}^4} e^{-4j\pi N_\lambda} e^{-j\pi} \\ &\times \int_x^\infty dx' e^{-2j\hat{k}(x'-\hat{x})} e^{-k_\omega (x'-\hat{x})} e^{-2(x'-\hat{x})^2/\sigma^2}. \end{aligned} \tag{A17}$$

Changing the integration variable

$$x' - \hat{x} = y, \tag{A18}$$

Eq. (A17) becomes

$$\begin{aligned} \psi_{\mathrm{nl}} \cong &-j\gamma \frac{\varepsilon_\gamma \omega^4 k_0^2 |b_+|^2 |\hat{\chi}_+|^4}{b_{\mathrm{nl}}^2 \omega_{\max}^4} e^{-4j\pi N_\lambda} e^{-j\pi} \\ &\times \int_{-\infty}^\infty dy e^{-2j\hat{k}y} e^{-k_\omega y} e^{-2y^2/\sigma^2}. \end{aligned} \tag{A19}$$

Changing again the integration variable

$$w = \frac{\sqrt{2}}{\sigma_x} y + \left(j\hat{k} + \frac{k_\omega}{2} \right) \frac{\sigma_x}{\sqrt{2}}, \tag{A20}$$

$$dw = \frac{\sqrt{2}}{\sigma_x} dy, \tag{A21}$$

$$\begin{aligned} \psi_{\mathrm{nl}} \cong &-j\gamma \frac{\varepsilon_\gamma \omega^4 k_0^2 |b_+|^2 |\hat{\chi}_+|^4}{b_{\mathrm{nl}}^2 \omega_{\max}^4} e^{-4j\pi N_\lambda} e^{-j\pi} \frac{\sigma_x}{\sqrt{2}} \\ &\times \int_{-\infty}^\infty dw e^{-w^2} \exp\left[\left(j\hat{k} + \frac{k_\omega}{2} \right)^{\wedge 2} \frac{\sigma_x^2}{2} \right], \end{aligned} \tag{A22}$$

which yields

$$\begin{aligned} \psi_{\mathrm{nl}} = &-j\gamma \frac{\varepsilon_\gamma \omega^4 k_0^2 |b_+|^2 |\hat{\chi}_+|^4}{b_{\mathrm{nl}}^2 \omega_{\max}^4} e^{-4j\pi N_\lambda} e^{-j\pi} \sigma_x \sqrt{\frac{\pi}{2}} \\ &\times \exp\left[\left(j\hat{k} + \frac{k_\omega}{2} \right)^2 \frac{\sigma_x^2}{2} \right], \end{aligned} \tag{A23}$$

or, separating the phase terms,

$$\begin{aligned} \psi_{\mathrm{nl}} \cong &-j\gamma \frac{\varepsilon_\gamma \omega^4 k_0^2 |b_+|^2 |\hat{\chi}_+|^4}{b_{\mathrm{nl}}^2 \omega_{\max}^4} e^{-4j\pi N_\lambda} e^{-j\pi} e^{j\hat{k} k_\omega \sigma_x^2/2} \sigma_x \sqrt{\frac{\pi}{2}} \\ &\times \exp\left[-\left(\hat{k}^2 - \frac{k_\omega^2}{4} \right) \frac{\sigma_x^2}{2} \right], \end{aligned} \tag{A24}$$

as in Eq. (29) of Talmadge *et al.* (2000).

The reflectivity function amplitude can be written as

$$\begin{aligned} |\psi_{\mathrm{nl}}| = &\frac{\varepsilon_\gamma}{2k(0) b_{\mathrm{nl}}^2} \frac{\omega^4 k_0^2}{\omega_{\max}^4} |b_+|^2 \eta^4 \left(\frac{p}{p_0} \right)^{4a} \frac{\hat{k}^6}{k(0)^6} \frac{\alpha}{k_\omega Q} \sqrt{\frac{\pi}{2}} \\ &\times \exp\left[-\left(\hat{k}^2 - \frac{k_\omega^2}{4} \right) \frac{\sigma_x^2}{2} \right], \end{aligned} \tag{A25}$$

which can be reduced, using Eq. (16), to

$$\begin{aligned} |\psi_{\mathrm{nl}}| \cong &\frac{\varepsilon_\gamma}{2b_{\mathrm{nl}}^2} \sqrt{\frac{\pi}{2}} \frac{k_0}{k_\omega} \frac{|a_+|^2}{\omega_{\max} \sigma_{\mathrm{BM}}^2} \left(\frac{p}{p_0} \right)^2 \eta^4 \left(\frac{p}{p_0} \right)^{4a} \frac{\alpha Q^2}{\omega^3} \\ &\times \exp\left[-\frac{\alpha^2 k_0^2}{2k_\omega^2 Q} \right]. \end{aligned} \tag{A26}$$

APPENDIX B: COCHLEAR REFLECTIVITY DUE TO ROUGHNESS

The perturbation is in this case associated with randomly distributed impedance fluctuations of the form

$$\tilde{\omega}_0(x) = \omega_0(x)(1 + r(x)\tilde{r}(x)), \tag{B1}$$

where

$$\omega_0(x) = \omega_{\max} e^{-k_\omega x}. \tag{B2}$$

As a consequence,

$$\tilde{\Delta}(\omega,x) = \tilde{\omega}_0^2(x) - \omega^2 + j\Gamma\omega = \omega_0^2(x) - \omega^2 + j\Gamma\omega + 2\omega_0^2(x)r\tilde{r}$$

$$= \Delta + 2\omega_0^2(x)r\tilde{r} = \Delta\left(1 + \frac{2\omega_0^2(x)r\tilde{r}}{\Delta}\right), \tag{B3}$$

$$\tilde{k}^2 = \frac{k_0^2\omega^2}{\tilde{\Delta}} = \frac{k_0^2\omega^2}{\Delta\left(1 + \frac{2r\tilde{r}\omega_0^2}{\Delta}\right)} \cong \frac{k_0^2\omega^2}{\Delta}\left(1 - \frac{2r\tilde{r}\omega_0^2}{\Delta}\right)$$

$$= k^2\left(1 - \frac{2r\tilde{r}\omega_0^2}{\Delta}\right). \tag{B4}$$

The perturbation to the wave vector is

$$\delta k_r^2 = k^2\frac{(-2r\tilde{r}\omega_0^2(x))}{\Delta} = k^2\frac{(-2r\tilde{r}\omega_0^2(x))}{k_0^2\omega^2}k^2. \tag{B5}$$

The reflectivity is given, as usual, by the expression

$$\psi_r \cong \gamma\int_x^\infty dx'\beta_+\delta k_r^2\beta_+. \tag{B6}$$

The usual approximations yield

$$\psi_r \cong -\gamma\int_x^\infty dx' k^2\beta_+\frac{2r\tilde{r}\omega_0^2(x')}{k_0^2\omega^2}k^2\beta_+$$

$$= -\gamma\frac{k_0^2\omega^2}{\omega_{\max}^4}\int_x^\infty dx' 2r\tilde{r}\omega_0^2(x')\chi_+^2. \tag{B7}$$

Developing the basis function

$$\chi_+(x,\omega) \approx \sqrt{\frac{k(0)}{\hat{k}}}\frac{\hat{k}^2}{k^2(0)}e^{-j\hat{k}(x-\hat{x})}e^{(x-\hat{x})^2/2\sigma_x^2}e^{-2j\pi N_\lambda}e^{-j\pi/2}$$

$$= \hat{\chi}_+ e^{-j\hat{k}(x-\hat{x})}e^{(x-\hat{x})^2/2\sigma_x^2}e^{-2j\pi N_\lambda}e^{-j\pi/2} \tag{B8}$$

and substituting it in Eq. (B7),

$$\psi_r = -\gamma\frac{\omega^2 k_0^2\hat{\chi}_+^2}{\omega_{\max}^4}e^{-4j\pi N_\lambda}e^{-j\pi}\int_x^\infty dx 2r\tilde{r}\omega_{\max}^2$$

$$\times e^{-2k_\omega\hat{x}'}e^{-2j\hat{k}(x'-\hat{x})}e^{-2k_\omega(x'-\hat{x})}e^{-(x'-\hat{x})^2/\sigma^2}. \tag{B9}$$

Substituting the roughness function with its FT

$$\tilde{r}(x) = \int_{-\infty}^{+\infty} dk e^{jkx}\tilde{r}(k), \tag{B10}$$

Eq. (B9) yields

$$\psi_r = -\gamma\frac{\omega^4 k_0^2\hat{\chi}_+^2 2r_0}{\omega_{\max}^4}e^{-4j\pi N_\lambda}e^{-j\pi}\int dk\int_x^\infty dx'\tilde{r}(k)$$

$$\times e^{jkx'}e^{-2j\hat{k}(x'-\hat{x})}e^{-2k_\omega(x'-\hat{x})}e^{-(x'-\hat{x})^2/\sigma^2}. \tag{B11}$$

The integral would give a negligible result if coherent filtering were not assumed. Coherent filtering picks up only $k=2\hat{k}$.

In Talmadge *et al.*, 1998, the following filtered roughness function is defined:

$$\tilde{r}_f(x) = \int_{-\infty}^{+\infty} dk_1 e^{jk_1x}\tilde{r}(k_1 + 2\hat{k})e^{-k_1^2\sigma^2/4}. \tag{B12}$$

This function is different from zero only around the value $k_1=0$ of the integration variable. We obtain

$$\psi_r = -\gamma\frac{\omega^4 k_0^2\hat{\chi}_+^2 2r_0}{\omega_{\max}^4}e^{-4j\pi N_\lambda}e^{-j\pi}\int_x^\infty dx'\tilde{r}_f(x')$$

$$\times e^{2j\hat{k}\hat{x}}e^{-2k_\omega(x'-\hat{x})}e^{-(x'-\hat{x})^2/\sigma^2}. \tag{B13}$$

Changing variables

$$x' - \hat{x} = y\sigma, \tag{B14}$$

$$\psi_r = -\gamma\frac{\omega^4 k_0^2\hat{\chi}_+^2 2r_0}{\omega_{\max}^4}e^{-4j\pi N_\lambda}e^{-j\pi}e^{2j\hat{k}\hat{x}}\sigma\int_{-\infty}^\infty dy\tilde{r}_f$$

$$\times(\hat{x} + y\sigma)e^{-2k_\omega y\sigma}e^{-y^2}, \tag{B15}$$

with a second variable change

$$\bar{y} = y + k_\omega\sigma, \tag{B16}$$

$$\psi_r = -\gamma\frac{\omega^4 k_0^2\hat{\chi}_+^2 2r_0}{\omega_{\max}^4}e^{-4j\pi N_\lambda}e^{-j\pi}e^{2j\hat{k}\hat{x}}e^{k_\omega^2\sigma_x^2}\sigma\int_{-\infty}^\infty d\bar{y}\tilde{r}_f$$

$$\times(\hat{x} + (\bar{y} - k_\omega\sigma)\sigma)e^{-\bar{y}^2}, \tag{B17}$$

$$\psi_r = -\gamma\frac{\omega^4 k_0^2\hat{\chi}_+^2 2r_0}{\omega_{\max}^4}e^{-4j\pi N_\lambda}e^{-j\pi}e^{2j\hat{k}\hat{x}}e^{k_\omega^2\sigma_x^2}\sigma\int dk_1\tilde{r}$$

$$\times(k_1 + 2\hat{k})e^{-k_1^2\sigma^2/4}e^{jk_1(\hat{x}-k_\omega\sigma^2)}\int_{-\infty}^\infty d\bar{y}e^{jk_1\sigma\bar{y}}e^{-\bar{y}^2}. \tag{B18}$$

The last variable change

$$\tilde{y} = \bar{y} - jk_1\sigma/2$$

yields

$$\psi_r = -\gamma\frac{\omega^4 k_0^2\hat{\chi}_+^2 2r_0}{\omega_{\max}^4}e^{-4j\pi N_\lambda}e^{-j\pi}e^{2j\hat{k}\hat{x}}e^{k_\omega^2\sigma_x^2}\sigma\int dk_1\tilde{r}$$

$$\times(k_1 + 2\hat{k})e^{-k_1^2\sigma^2/4}e^{jk_1(\hat{x}-k_\omega\sigma^2)}\int_{-\infty}^\infty d\tilde{y}e^{-k_1^2\sigma^2/4}e^{-\tilde{y}^2}, \tag{B19}$$

which gives

$$\psi_r = -\gamma\frac{\omega^4 k_0^2 \hat{\chi}_+^2 2r_0}{\omega_{\max}^4} e^{-4j\pi N_\lambda} e^{-j\pi} e^{2jk\hat{x}} e^{k_\omega^2\sigma_x^2}\sigma\sqrt{\pi}\int dk_1\tilde{r} \times (k_1 + 2\hat{k}) e^{-k_1^2\sigma^2/2} e^{jk_1(\hat{x}-k_\omega\sigma^2)}. \quad \text{(B20)}$$

Using the definition given in Eq. (B2),

$$\psi_r = -\gamma\frac{\omega^4 k_0^2 \hat{\chi}_+^2 2r_0}{\omega_{\max}^4} e^{-4j\pi N_\lambda} e^{-j\pi} e^{2jk\hat{x}} e^{k_\omega^2\sigma_x^2}\sigma_x\sqrt{\pi}\tilde{r}_f(\hat{x} - k_\omega\sigma_x^2), \quad \text{(B21)}$$

which can be reduced to

$$\psi_r = -\sqrt{\pi}\alpha r_0 \tilde{r}_f \frac{\omega^3 k_0}{\omega_{\max}^3 k_\omega Q}\hat{\chi}_+^2 e^{-4j\pi N_\lambda} e^{-j\pi} e^{2jk\hat{x}} e^{k_\omega^2\sigma_x^2}. \quad \text{(B22)}$$

Using also Eq. (15), it can be written as

$$\psi_r = -\sqrt{\pi}\alpha r_0 \tilde{r}_f \frac{k_0}{k_\omega}\eta^2\left(\frac{p}{p_0}\right)^{2a} Q^{1/2} e^{-4j\pi N_\lambda} e^{-j\pi} e^{2jk\hat{x}} e^{\alpha^2/Q^2}. \quad \text{(B23)}$$

APPENDIX C: REFLECTIVITY FROM SCALE-INVARIANCE VIOLATION

We assume a small violation of the scale-invariance symmetry

$$\Gamma(\omega,x) = \gamma_0 e^{-k_\gamma x} = \gamma_0 e^{-(k_\gamma - k_\omega)x} e^{-k_\omega x} = \gamma_0 e^{-k_\omega x}(1 - (k_\gamma - k_\omega)x) = \gamma_0(x) + \gamma_2(x),$$

$$(k_\gamma - k_\omega)x = dx \ll 1,$$

$$\gamma_2(x) = -dx\gamma_0(x) = -dx\frac{\gamma_0}{\omega_{\max}}\omega_{\max} e^{-k_\omega x} = -dx\varepsilon_\gamma\omega_0(x) \quad \text{(C2)}$$

Consequently,

$$\tilde{\Delta}(x) = \omega^2(x) - \omega^2 + j\Gamma\omega = \omega^2(x) - \omega^2 + j\gamma_0\omega + j\omega\gamma_2 = \Delta + j\omega\gamma_2\xi^2 = \Delta\left(1 + \frac{j\omega\gamma_2}{\Delta}\right), \quad \text{(C3)}$$

$$\tilde{k}^2 = \frac{k_0^2\omega^2}{\tilde{\Delta}} = \frac{k_0^2\omega^2}{\Delta\left(1 + \frac{j\omega\gamma_2}{\Delta}\right)} \cong \frac{k_0^2\omega^2}{\Delta}\left(1 - \frac{j\omega\gamma_2}{\Delta}\right) = k^2\left(1 - \frac{j\omega\gamma_2}{\Delta}\right), \quad \text{(C4)}$$

$$\delta k_{\text{siv}}^2 = k^2\frac{(-j\omega\gamma_2)}{\Delta} = k^2 dx\varepsilon_\gamma\frac{(+j\omega\omega_0(x))}{\Delta} = k^2 dx\varepsilon_\gamma\frac{(j\omega\omega_0(x))}{k_0^2\omega^2}k^2. \quad \text{(C5)}$$

The reflectivity is given by the equation

$$\psi_{\text{siv}} \cong \gamma\int_x^\infty dx'\beta_+\delta k_{\text{siv}}^2\beta_+. \quad \text{(C6)}$$

Proceeding as in the other appendixes,

$$\psi_{\text{siv}} = j\gamma\frac{d\varepsilon_\gamma\omega^4 k_0^2}{\Delta(0)^2}\int_x^\infty dx'\chi_+^2(x',\omega)e^{-k_\omega(x'-\hat{x})}x'. \quad \text{(C7)}$$

Developing the basis function

$$\psi_{\text{siv}} = j\gamma\frac{d\varepsilon_\gamma\omega^4 k_0^2|\hat{\chi}_+|^2}{\Delta(0)^2}e^{-4j\pi N_\lambda}e^{-j\pi}\int_x^\infty dx'x' \times e^{-2jk(x'-\hat{x})}e^{-k_\omega(x'-\hat{x})}e^{-(x'-\hat{x})^2/\sigma^2} \quad \text{(C8)}$$

and changing the integration variable

$$x' - \hat{x} = y,$$

Eq. (C8) becomes

$$\psi_{\text{siv}} \cong j\gamma\frac{d\varepsilon_\gamma\omega^4 k_0^2|\hat{\chi}_+|^2}{\Delta(0)^2}e^{-4j\pi N_\lambda}e^{-j\pi}\int_{-\infty}^\infty dy(\hat{x}+y) \times e^{-2jky}e^{-k_\omega y}e^{-y^2/\sigma^2}, \quad \text{(C9)}$$

with a new variable change

$$w = \frac{1}{\sigma_x}y + \left(jk + \frac{k_\omega}{2}\right)\sigma_x, \quad \text{(C10)}$$

$$dw = \frac{1}{\sigma_x}dy, \quad \text{(C11)}$$

$$\psi_{\text{siv}} \cong j\gamma\frac{d\varepsilon_\gamma\omega^4 k_0^2|\hat{\chi}_+|^2}{\Delta(0)^2}e^{-4j\pi N_\lambda}e^{-j\pi}\sigma_x\int_{-\infty}^\infty dwe^{-w^2} \times\left(\hat{x} + \sigma_x w - \left(jk + \frac{k_\omega}{2}\right)\sigma_x^2\right)\exp\left[\sigma_x^2\left(jk + \frac{k_\omega}{2}\right)^2\right] = j\gamma\frac{d\varepsilon_\gamma\omega^4 k_0^2|\hat{\chi}_+|^2}{\Delta(0)^2}e^{-4j\pi N_\lambda}e^{-j\pi}\sigma_x\left(\hat{x} - \left(jk + \frac{k_\omega}{2}\right)\sigma_x^2\right)\sqrt{\pi} \times\exp\left[\left(-k^2 + \frac{k_\omega^2}{4} + 2jk\frac{k_\omega}{2}\right)\sigma_x^2\right]. \quad \text{(C12)}$$

The reflectivity function amplitude can be written as

$$|\psi_{\text{siv}}| = \frac{d\varepsilon_\gamma\sqrt{\pi}}{2}\eta^2\left(\frac{p}{p_0}\right)^{2a}\frac{\hat{k}^3}{k_0^2}\frac{\alpha}{k_\omega Q}\sqrt{\left(\hat{x} - \sigma_x^2\frac{k_\omega}{2}\right)^2 + \hat{k}^2\sigma_x^4} \times\exp\left[-\sigma_x^2\left(\hat{k}^2 - \frac{k_\omega^2}{4}\right)\right]. \quad \text{(C13)}$$

Using the definitions of $\hat{k}$ and σ_x,

$$|\psi_{\text{siv}}| \cong d\varepsilon_\gamma\frac{\sqrt{\pi}k_0}{2k_\omega}\eta^2\left(\frac{p}{p_0}\right)^{2a} \times\sqrt{\left[\left(\hat{x} - \frac{\alpha^2}{2k_\omega Q^2}\right)^2 + \frac{\alpha^4 k_0^2}{k_\omega^4 Q^3}\right]}\alpha Q^{1/2} \times\exp\left[-\frac{\alpha^2 k_0^2}{k_\omega^2 Q}\right]. \quad \text{(C14)}$$

de Boer, E., and MacKay, R. (**1980**). "Reflections on reflections," J. Acoust. Soc. Am. **67**, 882–890.

Choi, Y., Lee, S., Parham, K., Neely, S. T., and Kim, D. O. (**2008**). "Stimulus-frequency otoacoustic emissions: Measurements in humans and simulations with an active cochlear model," J. Acoust. Soc. Am. **123**, 2651–2669.

Furst, M., and Lapid, M. (**1988**). "A cochlear model for acoustic emissions," J. Acoust. Soc. Am. **84**, 222–229.

Greenwood, D. D. (**1990**). "A cochlear frequency position function for several species: 29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hine, J. E., Ho, C. T., Slaven, A., and Thornton, A. R. (**2001**). "Comparison of transient evoked otoacoustic emission thresholds recorded conventionally and using maximum length sequences," Hear. Res. **156**, 104–114.

Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (**2004**). "Identification of otoacoustic emission components by means of adaptive approximations," J. Acoust. Soc. Am. **115**, 2148–2158.

Kalluri, R., and Shera, C. A. (**2007**). "Near equivalence of human click-evoked and stimulus-frequency otoacoustic emissions," J. Acoust. Soc. Am. **121**, 2097–2110.

Moleti, A., and Sisto, R. (**2008**). "Comparison between otoacoustic and auditory brainstem response latencies supports slow backward propagation of otoacoustic emissions," J. Acoust. Soc. Am. **123**, 1495–1503.

Notaro, G., Al-Maamury, A. M., Moleti, A., and Sisto, R. (**2007**). "Wavelet and Matching Pursuit estimates of the transient-evoked otoacoustic emission latency," J. Acoust. Soc. Am. **122**, 3576–3585.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.

Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.

Shera, C. A., and Guinan, J. J., Jr. (**1999**). "Evoked otoacoustic emissions arise from two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.

Shera, C. A., Tubis, A., and Talmadge, C. L. (**2005**). "Coherent reflection in a two-dimensional cochlea: Short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," J. Acoust. Soc. Am. **118**, 287–313.

Shera, C. A., Tubis, A., Talmadge, C. L., de Boer, E., Fahey, P. F., and Guinan, J. J., Jr. (**2007**). "Allen-Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions," J. Acoust. Soc. Am. **121**, 1564–1575.

Shera, C. A., and Zweig, G. (**1991**). "Reflection of retrograde waves within the cochlea and at the stapes," J. Acoust. Soc. Am. **89**, 1290–1305.

Sisto, R., and Moleti, A. (**2005**). "On the large scale spectral structure of otoacoustic emissions," J. Acoust. Soc. Am. **117**, 1234–1240.

Sisto, R., and Moleti, A. (**2007**). "Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels," J. Acoust. Soc. Am. **122**, 2183–2190.

Sisto, R., Moleti, A., and Shera, C. A. (**2007**). "Cochlear reflectivity in transmission-line models and otoacoustic emission characteristic time delays," J. Acoust. Soc. Am. **122**, 3554–3561.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (**2000**). "Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure," J. Acoust. Soc. Am. **108**, 2911–2932.

Tognola, G., Ravazzani, P., and Grandori, F. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.

Withnell, R. H., and McKinley, S. (**2005**). "Delay dependence for the origin of the nonlinear derived transient evoked otoacoustic emission," J. Acoust. Soc. Am. **117**, 281–291

Withnell, R. H., Hazlewood, C., and Knowlton, A. (**2008**). "Reconciling the origin of the transient evoked otoacoustic emission in humans," J. Acoust. Soc. Am. **123**, 212–221.

Zweig, G. (**1991**). "Finding the impedance of the organ of Corti," J. Acoust. Soc. Am. **89**, 1229–1254.

Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Effects of low-frequency biasing on spontaneous otoacoustic emissions: Frequency modulation

Lin Bian[a)]
Auditory Physiology Laboratory, 3430 Coor Hall, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

(Received 2 June 2008; revised 29 August 2008; accepted 2 September 2008)

It was previously reported that low-frequency biasing of cochlear structures can suppress and modulate the amplitudes of spontaneous otoacoustic emissions (SOAEs) in humans [Bian, L. and Watts, K. L. (2008). "Effects of low-frequency biasing on spontaneous otoacoustic emissions: Amplitude modulation," J. Acoust. Soc. Am. **123**, 887–898]. In addition to amplitude modulation, the bias tone produced an upward shift of the SOAE frequency and a frequency modulation. These frequency effects usually occurred prior to significant modifications of SOAE amplitudes and were dependent on the relative strength of the bias tone and a particular SOAE. The overall SOAE frequency shifts were usually less than 2%. A quasistatic modulation pattern showed that biasing in either positive or negative pressure direction increased SOAE frequency. The instantaneous SOAE frequency revealed a "W-shaped" modulation pattern within one biasing cycle. The SOAE frequency was maximal at the biasing extremes and minimized at the zero crossings of the bias tone. The temporal modulation of SOAE frequency occurred with a short delay. These static and dynamic effects indicate that modifications of the mechanical properties of the cochlear transducer could underlie the frequency shift and modulation. These biasing effects are consistent with the suppression and modulation of SOAE amplitude due to shifting of the cochlear transducer operating point. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2990716]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.64.Bt, 43.64.Yp [BLM] Pages: 3009–3021

I. INTRODUCTION

Our inner ears are not only sound receivers but also sound generators. Even without an external acoustic stimulus, the inner ear can produce sounds that are called spontaneous otoacoustic emissions (SOAEs). These internally generated sounds are commonly identified by spectral analysis as tonelike peaks with frequencies distributed in the range from 0.5 to 5 kHz (see Probst *et al.*, 1991 for a review). On average, the number of the SOAE components is about 6 in each ear, but can be as many as 32/ear depending on the criterion for the detection of spectral peaks (Talmadge *et al.*, 1993). The amplitudes of these sounds are usually below 10 dB sound pressure level (SPL), thus rarely audible. If detection techniques are sensitive enough, SOAEs can be identified in up to 90% of normal hearing individuals (Pasanen and McFadden, 2000). The number of SOAE components, the distribution of SOAE frequencies, and their relative amplitudes are unique and relatively stable just like a "fingerprint" of a given inner ear.

Despite the relative stability and individual uniqueness, there are some remarkable differences in the features of SOAEs between genders and laterality. Females generally show more SOAE components with larger amplitudes than males (Bilger *et al.*, 1990). This gender difference may be a genetic trait or a consequence of differential influence of sexual hormones on embryonic neural development (McFadden, 2002). The prevalence of SOAE in the right ear is usually higher than the left (Kuroda, 2007). Such an asymmetry seems to indicate a correlation with the differential specialization of the central auditory system (Khalfa and Collet, 1996) or at least an influence of the efferent system (Mott *et al.*, 1989). These top-down connections between the brain and SOAEs in the inner ear suggest a possible involvement of outer hair cells (OHCs) in the formation of SOAEs because only these cells are largely innervated by efferent neural fibers from the brainstem. It is known that the OHCs are mechanically active due to their voltage-dependent motile responses in the cellular membrane (see Ashmore, 2008 for a review) or in the hair bundles (Martin *et al.*, 2003). The disinhibition due to irregular and sparse efferent innervations in the apical regions of the cochlea (Pujol, 2001; Thiers *et al.*, 2002) could allow such mechanical vibrations from the OHCs at certain locations to build up and eventually escape the boundary of the inner ear. The SOAEs may be carried out of the cochlea and stabilized by standing waves formed between the stapes footplate and these OHCs or other vibrating structures (Shera, 2003). The apical wave reflectors are scattered largely near the best frequency location of the SOAE. Hence, the OHC is a crucial element in the generation of SOAEs.

The essential function of the hair cells is to transform mechanical energy in sound into neural pulses through a transduction process which is known to be nonlinear. The transduction in OHC is bidirectional where mechanical vibrations from the reverse process could boost hearing sensitivity. The nonlinearity is responsible for a variety of auditory phenomena, such as compressive growth of cochlear responses, two-tone suppression, and generation of distortion

[a)]Author to whom correspondence should be addressed. Electronic mail: lin.bian@asu.edu

product (Robles and Ruggero, 2001). These nonlinear features of OHC transducers are reflected in the behaviors of SOAEs to a range of experimental manipulations of cochlear mechanics. Presenting an external tone can suppress SOAEs and generate distortion products (Rabinowitz and Widin, 1984; Long *et al.*, 1993; Norrix and Glattke, 1996). A suppression tuning curve can be constructed since the suppressor level has to be progressively reduced when its frequency approaches the SOAE (Bargones and Burns, 1988). It has been noticed that distortion product otoacoustic emissions (DPOAEs) elicited by two tones are larger in ears that generate SOAEs than those do not (Ozturan and Oysu, 1999; Moulin, 2000). This observation indicates that the two different types of emissions could share a common element in their generating mechanisms which could be the nonlinear hair cell transduction.

Given that the cochlear nonlinearity is essential for normal hearing sensitivity and frequency selectivity, estimating the characteristics of the cochlear transducer has significant clinical utility in differential diagnosis of inner ear disorders. Toward this goal, a low-frequency biasing technique has been used to assess the dynamical nonlinearity of the inner ear (Bian *et al.*, 2002, 2004; Bian and Scherrer, 2007). The bias tone can shift the cochlear partition and produce an amplitude modulation (AM) of cochlear distortion products. The result shows that the amplitude change of the DPOAEs by the bias tone is consistent with the effect of a sigmoid-shaped transducer function. Since the cochlear transducer nonlinearity is sensitive to physiological and pathological alternations in the inner ear, it is hypothesized that mechanically varying the cochlear transducer gain could influence SOAEs. To test the hypothesis, the low-frequency biasing technique was applied to human subjects with large SOAEs (Bian and Watts, 2008). The results showed that SOAE amplitudes were suppressed and temporally modulated by the bias tone. The pattern of the SOAE magnitude variation resembled the first derivative of the sigmoid-shaped cochlear transducer function. In addition, it was noticed that the SOAE frequencies were changed by the bias tone. The frequency of SOAE is another sensitive indicator of alterations in cochlear mechanics since it can be influenced by body temperature (O'Brien, 1994), intracochlear pressure (de Kleine *et al.*, 2000), external sound (Long *et al.*, 1991, 1993), and activation of the efferent system (Mott *et al.*, 1989). This paper investigates the frequency shift and frequency modulation (FM) of SOAEs under low-frequency biasing.

II. METHODS

A. Experimental procedures

The experimental procedures were described in the earlier report (Bian and Watts, 2008). Briefly, 11 normal hearing ears were selected for having at least one large SOAE which was defined by a greater than 20 dB signal-to-noise ratio (SNR). The ear-canal acoustics were recorded with a calibrated probe microphone (Etymōtic Research, ER-10B+) while presenting a low-frequency bias tone with an insertion earphone (HA-FX55, JVC). The bias tone was digitally generated and controlled with software implemented in LABVIEW [Version 8, National Instruments (NI)]. The bias tone was delivered from a 24 bit dynamic signal acquisition and generation card (PXI-4461, NI) at 25, 32, 50, 75, and 100 Hz. To result in a 1 Hz frequency resolution in subsequent spectral analyses, the duration of the bias tone was set to 1 s. At each bias tone frequency (f_{bias}), the peak amplitude was attenuated automatically from 3 or 4 Pa to 0 in 41 steps. The maximal biasing pressure used was determined by increasing the bias tone level (L_{bias}) until a significant amplitude reduction in the largest SOAE was observed. At each biasing step, the recorded ear-canal signal was amplified 20 dB, averaged up to 16 times, and digitized at a sampling frequency (f_{samp}) of 204.8 kHz. This procedure was repeated with a 20 min break to measure the variability and repeatability of the biasing effects between the two trials.

B. Signal processing and data analysis

1. General spectral analysis

Data were analyzed off-line in MATLAB (Version 7, MathWorks). A general spectral analysis was applied to the signals collected at different L_{bias} to extract the features of the SOAEs. First, the entire signal was high pass filtered at 400 Hz to eliminate the bias tone and submitted to a fast Fourier transform (FFT) to obtain a set of 41 spectra. Second, the SOAE components were identified as spectral peaks with 20 dB SNR in the absence of a bias tone, and their baseline frequencies and amplitudes were recorded. Third, SOAE frequencies were extracted across all records. For conditions when a SOAE was suppressed, the SOAE frequency (f_{SOAE}) was determined by the largest spectral peak within $\pm 3f_{bias}$ from the baseline value. Finally, the amplitudes of sidebands above and below the SOAE component or the upper and lower sidebands (USBs and LSBs) were measured. Spectral sidebands, typical for both AM and FM signals, were located with frequency intervals of integer multiples of the f_{bias} from the SOAE. Two LSBs and USBs, namely, sidebands I and II that are f_{bias} and $2f_{bias}$ away from the SOAE [Fig. 1(A)], were used in further analyses.

2. Quasistatic modulation pattern

To examine the influence of biasing directions on f_{SOAE}, a quasistatic modulation pattern (Bian *et al.*, 2002; Bian and Chertoff, 2006) was obtained with analyzing windows fixed at the peaks and troughs of the bias tones [Fig. 1(B)]. The window length (l_{win}) was limited within one-half biasing cycle to ensure that the biasing effect on SOAEs was due to a "static" displacement of the cochlear partition in the same direction. Detecting small frequency changes in SOAE required high-resolution spectral analyses, thus a short-time FFT was insufficient. Alternatively, these short-analyzing windows were linked [Fig. 1(B)] to yield a long time sequence, thus a significant reduction in frequency resolution could be avoided. Due to the periodicity of the repeating windows, a consequence of this linked-window method was the production of harmonic peaks that could interfere with the measurement of SOAE components. This could be largely eliminated by adjusting the l_{win} in a few trials. Ini-

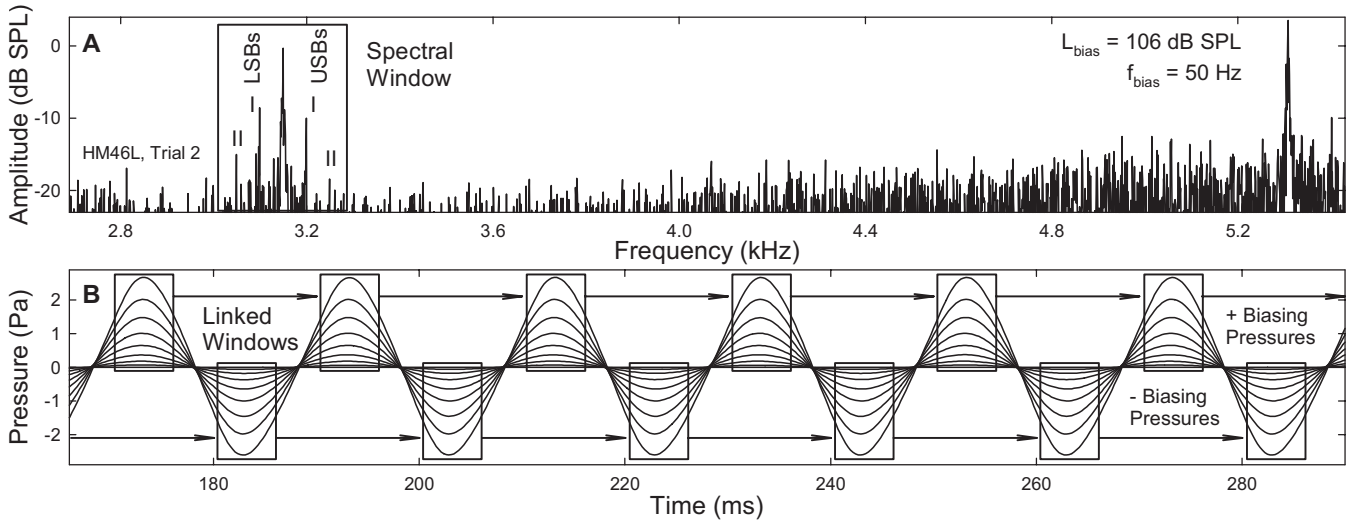


FIG. 1. Analysis methods. (A) A spectrum of the ear-canal signal and the spectral window method. The box represents a rectangular window centered at a SOAE containing two sidebands on each side. The windowed contents are submitted to an IFFT to extract the temporal features of the SOAE. LSBs: lower sidebands; USBs: upper sidebands. Sidebands I: frequency spacing with the SOAE equals f_{bias}; sidebands II: frequency spacing with the SOAE equals $2f_{bias}$. (B) Linked-window method. Temporal windows located at the peaks and troughs of a sequence of bias tones with various amplitudes are linked, respectively, for spectral analysis. The calculated SOAE amplitudes and frequencies represent the two halves of the quasistatic modulation pattern in the positive and negative sound pressure directions, respectively.

tially, six different l_{win} were selected from a set of values ranging between $\frac{1}{4}$ and $\frac{1}{2}$ biasing cycle with a $\frac{1}{16}$ cycle increment, i.e., $l_{win}=(n/16)N_{bias}$, where N_{bias} is the number of points within a biasing cycle given by f_{samp}/f_{bias} and $n=3,4,\ldots,8$. For each l_{win}, the frequency difference between the f_{SOAE} and the closest harmonic of f_{bias} was calculated. The l_{win} with the largest frequency difference was selected, and the analyzing windows were linked for FFT. If the frequency of SOAE obtained with the linked-window method (f'_{SOAE}) from the no-biasing condition drifted from the baseline f_{SOAE}, then the frequency ratio (f'_{SOAE}/f_{SOAE}) was used as a factor to further adjust the window length:

$$l_{win}=\frac{n}{16}\frac{f'_{SOAE}}{f_{SOAE}}N_{bias}. \quad (1)$$

This adjustment could ensure accuracy in f_{SOAE} estimation and eliminate the harmonics generated by linked windows. Then, the new time sequences were zero padded on both ends to result in a 1 Hz frequency resolution in FFT. Finally, the SOAE frequencies obtained for the negative and positive biasing pressures were combined to form a quasistatic modulation pattern.

3. Period modulation pattern

A spectral windowing method was employed to extract the temporal pattern of the SOAE frequency variation [Fig. 1(A)]. The positive half of the complex spectrum of the ear-canal signal was selected by a rectangular window centered at the f_{SOAE}. Initial analysis indicated that the SOAE frequency change (Δf_{SOAE}) was small compared to the f_{bias}, and only the two sidebands closest to the SOAE in both LSB and USB contained most energy. Therefore, a narrower bandwidth of the spectral window than the AM analysis (Bian, 2006; Bian and Watts, 2008) was determined (see Appendix and Fig. 12 for details):

$$B_{FM}=2(\Delta f_{SOAE}+2f_{bias}). \quad (2)$$

The windowed complex spectrum was submitted to an inverse fast Fourier transform (IFFT) to obtain an analytic signal of the frequency modulated SOAE. Its instantaneous frequency (f_i) was calculated by the derivative of the unwrapped phase (φ) with respect to time:

$$f_i(t)=\frac{f_{samp}}{2}\frac{1}{2\pi}\frac{d\varphi}{dt}, \quad (3)$$

where $f_i(t)$ is a time waveform representing the f_{SOAE}. A two-point average smoothing was applied to the phase derivative to reduce the roughness or variability, and the waveform was down sampled to $\frac{1}{20}$ of the f_{samp} for further use. To establish the correlation between the temporal features of f_{SOAE} with the bias tone phase, the f_{SOAE} waveform was truncated based on each of the 22–88 biasing cycles and averaged to reduce noise. This resulted in a period modulation pattern of the f_{SOAE} for each biasing step.

III. RESULTS

A. Spectral fine structures

1. Upward frequency shift

Since the general characteristics of SOAEs and the biasing effects on SOAE amplitudes in these ears were described in a preceding report (Bian and Watts, 2008), the present paper focuses on the frequency change induced by the bias tone. The frequencies and magnitudes of SOAEs were highly dependent on each individual ear; thus the results will be reported using examples and efforts will be made to reflect the general trends in all observations. To examine the frequency change, the variability due to random error in frequency analysis and normal frequency fluctuation over time were assessed by checking repeated measures of the baseline f_{SOAE} at the last biasing step ($L_{bias}=0$). Averaged

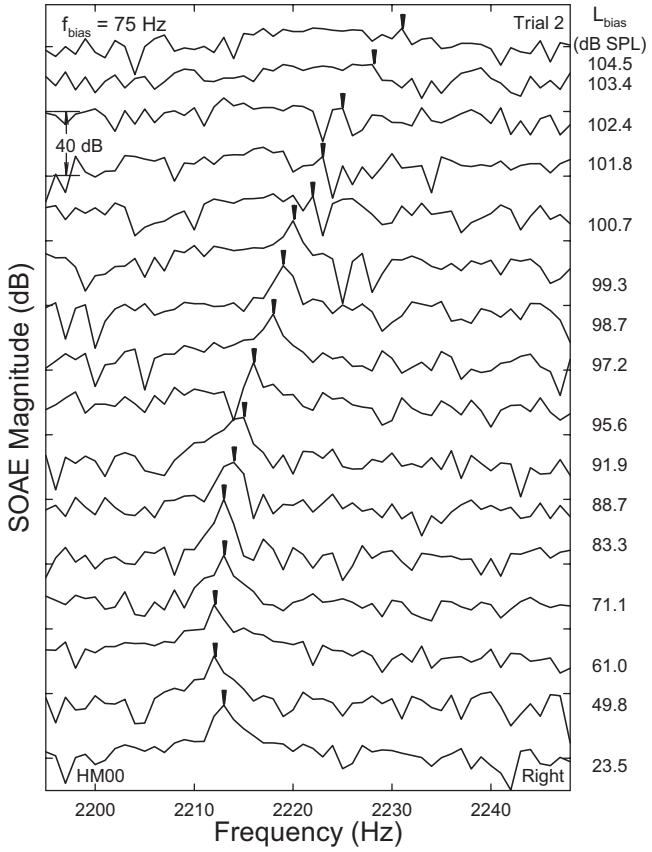


FIG. 2. Upward frequency shift of SOAE with L_{bias}. A sequence of spectra focused on a SOAE with an increase in the L_{bias} demonstrates an upward-frequency shift of the SOAE. Each spectrum has been shifted 40 dB with respect to the next to avoid overlap. Note that some spectra without obvious frequency shift at lower L_{bias} are omitted.

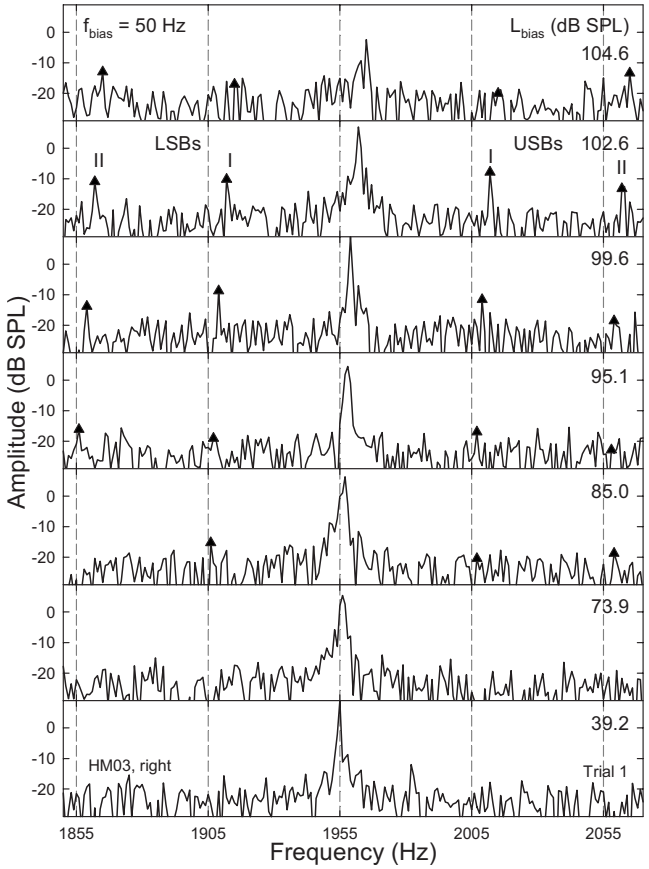


FIG. 3. Spectral effects of L_{bias} on a SOAE: frequency shift, generation of sidebands, and suppression. Spectra shown are obtained at moderate to high bias tone levels where these effects are usually observed. The sidebands (▲) are presented at integer multiples of the f_{bias} from the SOAE, thus shifting with the SOAE. Note that the sidebands initially increase with the L_{bias} and then fall as the SOAE is suppressed. Sidebands I and II: frequency spacing with f_{SOAE} equals f_{bias} and $2f_{bias}$, respectively.

across all ears and bias tone frequencies, the standard deviations for trials 1 and 2 were about 2 and 1.5 Hz, respectively. The change in the baseline f_{SOAE} between the two trials averaged across all emission components in all ears was about +1 Hz. However, this increase in baseline f_{SOAE} was within the 2 Hz standard deviation of the baseline f_{SOAE} estimated across trials. Unless otherwise noted, the SOAE frequencies of the two trials were averaged. This analysis indicated that any SOAE frequency shift, if more than 2 Hz, could be considered as significant change.

The most important findings of the study were an upward shift of the f_{SOAE} and a suppression of the SOAE magnitude due to the presence of the bias tone. Even though the amount of frequency shift was commonly less than 20–40 Hz, the frequency shift was certainly noticeable when comparing spectra obtained at low versus high bias tone levels (Fig. 2). The direction of the frequency shift was always an increase which occurred in most of the SOAEs from all ears. These upward-frequency shifts could readily be observed under moderate to high bias tone levels, usually between around 70 and 100 dB SPL where there was no significant SOAE amplitude reduction. When the SOAE amplitude started to be suppressed, the frequency shift became more dramatic, sometimes accompanied by a widening of the spectral peak. At lower bias tone levels (<65–85 dB SPL), only a fluctuation of the f_{SOAE} could be observed while the SOAE amplitude remained unaffected.

2. Spectral sidebands

Another frequency-domain observation was the presence of multiple sidebands around the SOAE when relatively high bias tone levels were used. In most cases, the sidebands were detected between 80 and 90 dB SPL L_{bias} where SOAE amplitudes were stable (Fig. 3). The number and sizes of the sidebands on either side of the SOAE varied depending on the bias tone strength relative to the SOAE (middle panels). Under strong influence of the bias tone, there were usually large multiple sidebands that could mostly be attributed to the AM of the SOAE (Bian and Watts, 2008). At very high biasing levels where the SOAEs were suppressed, the sizes of the sidebands also decreased (top panel). As the L_{bias} became weaker (<90–100 dB SPL), one or two smaller sidebands appeared on each side of the SOAE. These sidebands could mainly come from the FM of the SOAE because the SOAE amplitude was stable at the maximal value. Usually, sidebands I in either LSB or USB were the largest in magnitude. Then the next prominent sidebands were sidebands II. More distant sidebands were generally very small or absent. When the L_{bias} was below 70 dB SPL, typically no sidebands were observed (bottom).

As the L_{bias} increased, the frequency and amplitude of SOAE demonstrated some opposite behaviors (Fig. 4). The

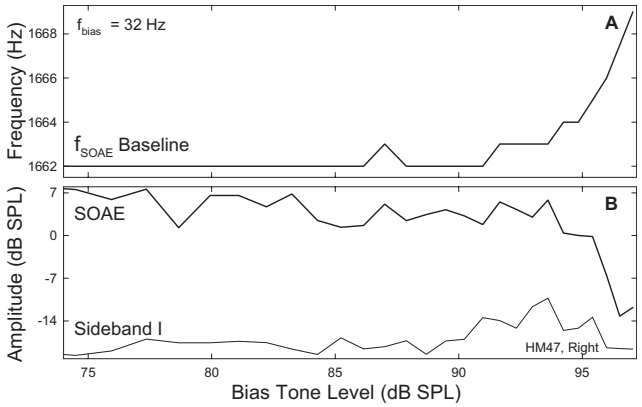


FIG. 4. Biasing effects on SOAE frequency, amplitude, and sidebands. (A) An upward-frequency shift of SOAE with the L_{bias}. (B) Suppression of the SOAE amplitude (top trace) and growth of sidebands (lower trace). The f_{SOAE} starts to increase before its amplitude is significantly reduced. The sideband becomes measurable at a much lower L_{bias} (75 dB SPL). Data are average of the two trials.

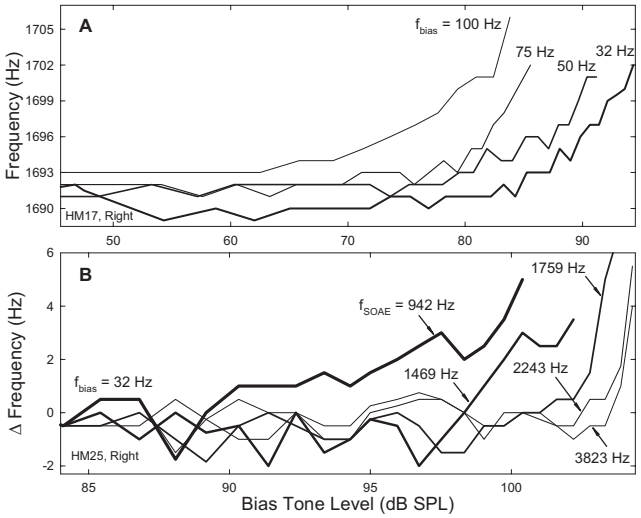


FIG. 5. Frequency shifts: differential effects of f_{bias} and f_{SOAE}. (A) A sequence of traces showing the frequency shift of a SOAE at 1691 Hz under the influence of bias tones with different frequencies. Note that the f_{SOAE} begins to increase at a lower L_{bias} and can reach a higher final maximal value as the f_{bias} increases. (B) Frequency shifts from the baseline values for a set of SOAEs under the influence of a 32 Hz bias tone. The bias tone is more effective for lower-frequency SOAEs since lower bias tone levels are required. Data are average of the two trials.

changes of sidebands were represented by the relatively large sidebands I in the LSB and USB (lower trace in panel B). When L_{bias} was below 85 dB SPL, the SOAE magnitude fluctuated around 5 dB SPL without any frequency change. When the L_{bias} exceeded 85 dB SPL, the f_{SOAE} started to rise, the sideband grew continuously, and the SOAE magnitude began to fall. Above 95 dB SPL, the SOAE was significantly suppressed and sideband showed a rollover (panel B), while the f_{SOAE} reached a maximal value (panel A). The growth of the sidebands became noticeable when the L_{bias} was over 75 dB SPL and was steeper above 88 dB SPL. From this L_{bias} to where the sidebands reached their peaks, both the SOAE frequency and amplitude were unstable and eventually proceeded to opposite changes.

3. Frequency effects

It was observed that the biasing effects on a SOAE were highly dependent on the relative frequency difference between the bias tone and the SOAE. The bias tone was more effective when the f_{bias} was closer to the f_{SOAE}. The influence of bias tones with different frequencies on the same SOAE reflected a range of such effectiveness [Fig. 5(A)]. The 100 Hz bias tone was the most effective since at only 65 dB SPL it could cause a frequency increase and a 13 Hz shift at 84 dB SPL. The 32 Hz bias tone was the weakest one because the f_{SOAE} remained relatively unchanged until the L_{bias} reached above 85 dB SPL and the frequency shift was smaller. SOAE frequency shifts for bias tones of 75 and 50 Hz were distributed in between. The effectiveness of a bias tone could also be evident by frequency shifts of multiple SOAEs [Fig. 5(B)]. It could be observed that the bias tone was much more effective in producing a frequency shift of a 942 Hz SOAE than other SOAEs with higher frequencies. For SOAEs of 1800 Hz or higher, the L_{bias} had to be greater than 100 dB SPL to produce any measurable frequency shift. This range of effects indicated that the bias tone was less effective when the SOAE frequency locations were farther away from the cochlear apex.

Data compiled from all the SOAEs under various biasing conditions showed such a reduction in biasing effect with f_{SOAE} (Fig. 6). Two quantities could be assessed from the data: the maximal Δf_{SOAE} and the L_{bias} required to produce a just noticeable frequency shift of 2 Hz. For a fixed f_{bias}, the maximal Δf_{SOAE} decreased as the f_{SOAE} increased (Fig. 6, top row). These results suggested that the influence of the bias tone diminished while the SOAE frequency place moved toward the cochlear base. Depending on the f_{bias}, the rate of reduction in biasing effects varied from roughly −6 to −16 Hz/octave increase in f_{SOAE}. For an ear with multiple SOAEs spreading from 1 to 5 kHz, the frequency shift was mostly observed below 3–4 kHz. Moreover, the Δf_{SOAE} increased mainly from about 10 to 30 Hz when the f_{bias} varied from 25 to 100 Hz. These frequency shifts amounted to approximately 1%–2% of the baseline f_{SOAE}. The second quantity, the L_{bias} needed to observe a minimal Δf_{SOAE}, increased with the f_{SOAE} (Fig. 6, bottom row). This indicated that one had to increase the L_{bias} or reduce the frequency difference between the bias tone and the SOAE to observe an effect. The rate of level increase seemed to be constant across all the bias tone frequencies. A rough estimate showed that the demanding of the biasing power increased at a rate of 8 dB/octave rise in f_{SOAE}. The L_{bias} for minimal Δf_{SOAE} ranged from 65 to 95 dB SPL for f_{bias} of 100 and 25 Hz, respectively.

B. Quasistatic modulation pattern

Quasistatic modulation patterns obtained from linked segments of the acoustic signal fixed at peaks and troughs of bias tones of various amplitudes provided a measure of f_{SOAE} variations as a result of biasing cochlear structures in opposite directions. A pair of exemplary f_{SOAE} modulation patterns from an ear showed similar characteristics across the

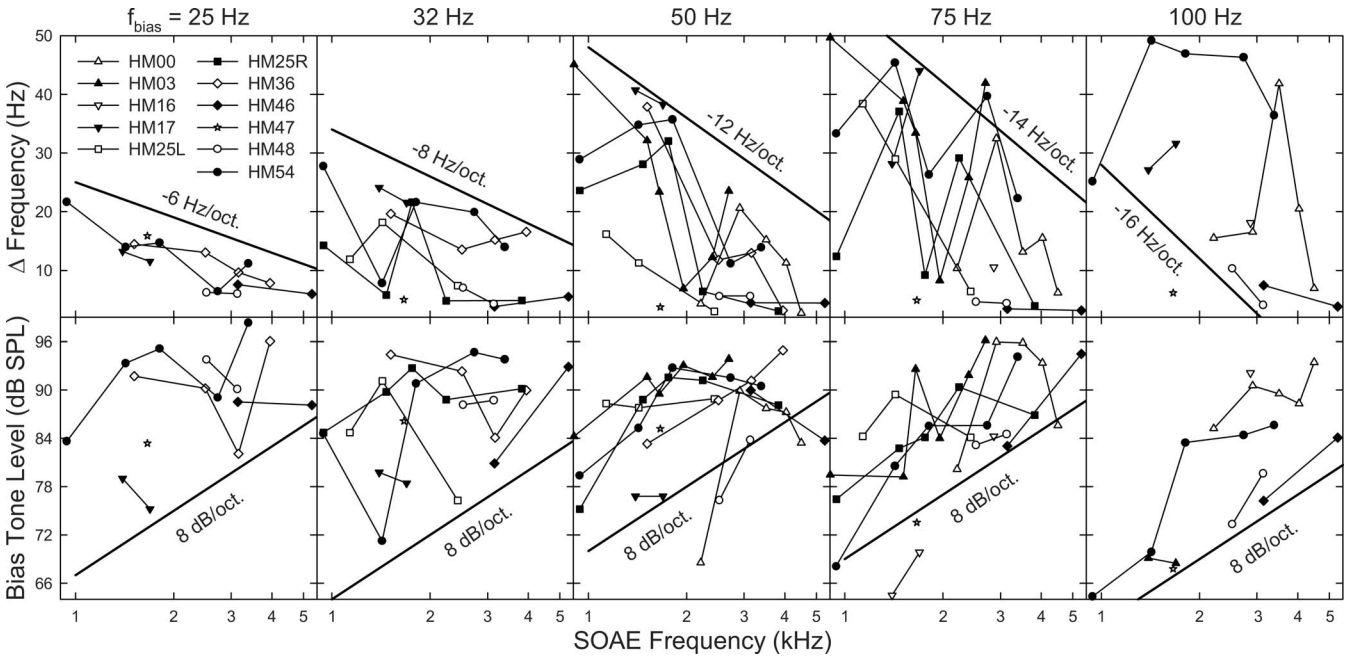


FIG. 6. Maximal SOAE frequency shift and effective bias tone levels. Top row: The maximal frequency shifts of all SOAEs under the influence of bias tones at different frequencies. The approximated reduction rate in the maximal frequency shift ranging from −6 to −16 Hz/octave is depicted by a thicker line in each panel. Bottom row: The bias tone levels required to produce a minimal frequency shift of 2 Hz for all the SOAEs are shown for different bias tone frequencies. Generally, it requires more sound pressure to produce a frequency shift for higher-frequency SOAEs. The rate of reduction in the bias tone efficiency is roughly 8 dB/octave (thicker lines) regardless of the f_{bias}. Data are average of the two trials.

two trials (Fig. 7). The f_{SOAE} was minimal around zero biasing pressure and shifted upward when the biasing pressure stepped up in either direction. The "U-shaped" modulation pattern expanded between ±3 Pa biasing pressures and revealed a modulation depth, the difference between the maximal and minimal SOAE frequencies, of about 15–20 Hz. The f_{SOAE} showed large fluctuations at the extremes of the biasing pressures. This increased variability was a consequence of suppressions of SOAE to the noise floor, leading to random errors in the frequency estimation. A smaller fluctuation of the f_{SOAE} in the amount of 3–4 Hz due to low biasing pressures could also be noted at the bottom of the modulation patterns.

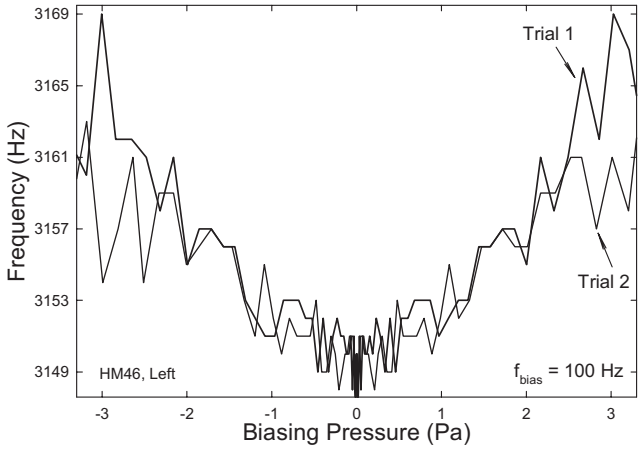


FIG. 7. Quasistatic modulation patterns of f_{SOAE}. Biasing in either positive or negative sound pressure directions increases the f_{SOAE}. Fluctuation of the f_{SOAE} about the baseline value (3148 Hz) can be observed around zero biasing pressure. Data obtained from the two trials with the linked-window method are consistent.

There were frequency-dependent variations in the quasistatic modulation patterns. This could be observed by examining the behavior of a single SOAE under different biasing frequencies (Fig. 8, left column). As the f_{bias} increased (from bottom up), the modulation depth increased and the width reduced. The modulation depth varied from about 20 to 45 Hz as the f_{bias} was elevated from 25 to 100 Hz. The modulation width, the range of biasing pressures between the maximal SOAE frequencies, expanded across ±1 Pa at 25 Hz biasing condition and gradually narrowed to only ±0.5 Pa when the f_{bias} was above 75 Hz. Larger frequency shifts at lower biasing pressures indicated that higher-frequency bias tones were more effective. The second frequency effect could be evident from a single ear as the modulation patterns of different SOAEs by the same bias tone (Fig. 8, right column). As the f_{SOAE} increased (from top down), the modulation patterns became wider and shallower, implying a weakening of the biasing effect. The SOAE magnitude and frequency also influenced the modulation pattern. Larger emissions (e.g., 1.95 kHz) showed a better defined pattern with less fluctuation. Lower-frequency SOAEs (e.g., 1.5 kHz) were easily suppressed by the bias tone, thus showing a shallower pattern. For larger SOAEs, more frequency shifts could be achieved but at higher biasing pressures.

C. Period modulation pattern

It should be noted that the SOAEs were nonstationary signals since their amplitudes and frequencies varied depending on the relative power of the external tone and the direction of mechanical biasing. It was necessary to analyze the temporal property of the f_{SOAE} in relation to the cyclic biasing of cochlear structures. The instantaneous f_{SOAE} wave-

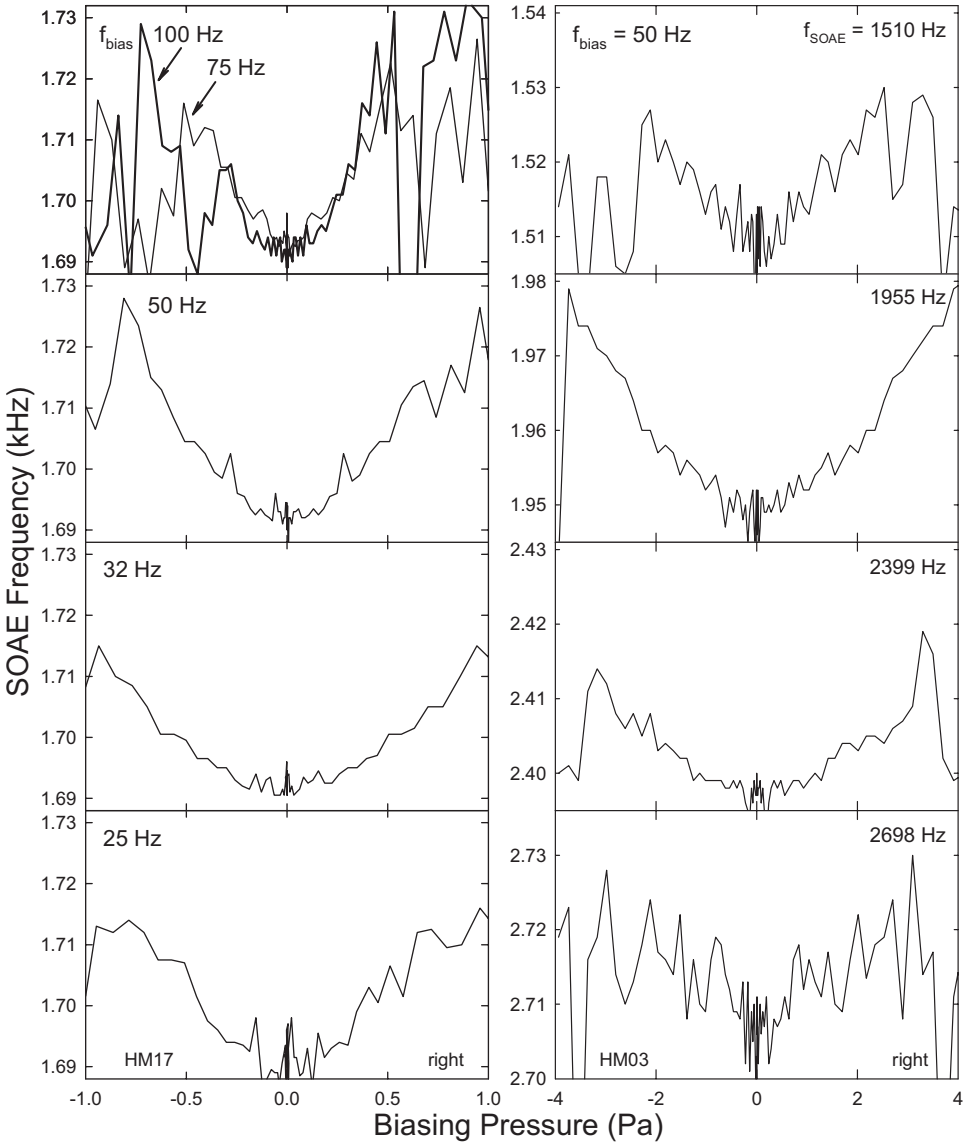


FIG. 8. Quasistatic modulation patterns: frequency effects. Left column: Effect of f_{bias} on a SOAE at 1691 Hz. From top to bottom, as f_{bias} decreases, the width of the pattern increases, and the modulation depth decreases. Note that at both biasing extremes, especially, for $f_{bias}>75$ Hz, the f_{SOAE} shows large variations due to significant suppressions of the SOAE. Right column: Effects of a 50 Hz bias tone on four SOAEs in an ear. From top to bottom, as f_{SOAE} increases, the modulation pattern becomes shallower and wider. The modulation pattern is also influenced by the SOAE amplitude, e.g., the best defined pattern is that of the largest SOAE (1955 Hz).

form obtained from IFFT demonstrated some periodicity and randomness [Fig. 9(A)]. When closely examined with respect to the corresponding bias tone, it could be found that there were three maxima and two minima in one biasing cycle. This regularity was verified by performing a FFT on the f_{SOAE} waveform [Fig. 9(B)]. A large spectral peak was revealed at the second harmonic of the bias tone or $2f_{bias}$, indicating that a pattern appeared in every half basing cycle. Another way of examining the dynamic frequency change was to study the distribution of frequency points by constructing a histogram over the entire waveform [Fig. 9(C)]. The histogram showed two features of the f_{SOAE}: peak splitting and exact bandwidth. Peak splitting implied that the f_{SOAE} oscillated between two values, the baseline and a maximum, similar to a sinusoid wave. The width of the histogram as obtained by the 5th and 95th percentiles of the cumulative distribution of frequencies provided a measure of the variability of f_{SOAE}, i.e., 90% of the frequencies occurred within this range. Such a range could be supplied to Eq. (2) to refine the filter bandwidth for analyzing the FM signal.

Because the bias tones could modulate both SOAE amplitude and frequency (Figs. 2–4), for a given f_{bias}, instantaneous frequencies obtained from moderate bias tone levels ($<$95–100 dB SPL) with no SOAE suppression were examined to determine the temporal modulation pattern. The instantaneous f_{SOAE} showed some large transient fluctuations [Fig. 9(A)] that could partially be related to random variations in SOAE amplitude. To further reduce the random errors in frequency estimation, a period modulation pattern was derived from averaging the instantaneous f_{SOAE} over a majority of the biasing cycles. Other periods where the instantaneous frequency exceeded the 90% confidence range were excluded from averaging. Only when more than half of the total biasing periods were included for averaging was a period modulation pattern considered valid [Fig. 9(D)]. The period modulation pattern typically showed a "W" shape: a center peak and two notches. Compared with the bias tone, frequency peaks occurred at the biasing extremes, and minima were close to zero crossings of the bias tone. Stronger SOAEs were more likely to show the typical pattern in more biasing cycles.

For a fixed f_{bias}, this typical modulation pattern varied with the L_{bias}. Two trends could be observed with the reduc-

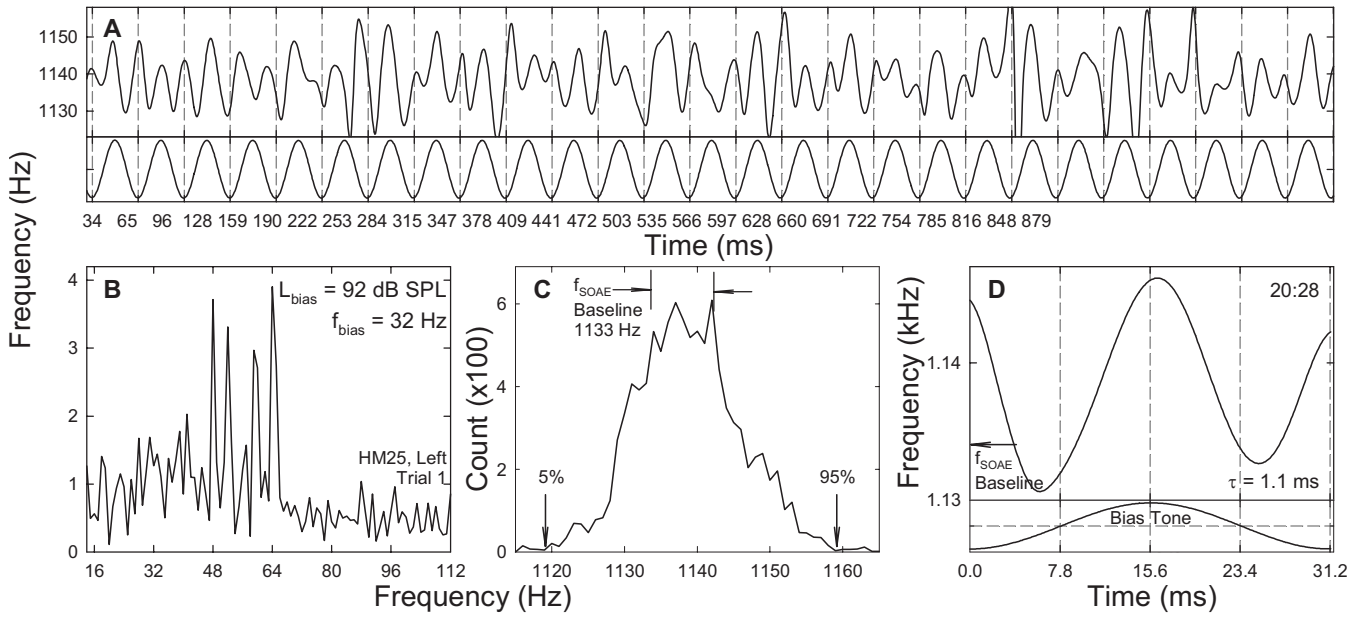


FIG. 9. Characteristics of temporal pattern of f_{SOAE}. (A) The instantaneous frequency of a SOAE in relation to the bias tone (lower trace). Compared with the bias tone cycles, it can be observed that the f_{SOAE} peaks three times in a biasing period. (B) A spectrum of the f_{SOAE} waveform in panel A. The most prominent spectral peak appears at 64 Hz or $2f_{bias}$, indicating that a pattern occurs twice in a biasing cycle. (C) A histogram of the oscillating f_{SOAE} shows peak splitting. The left branch of the peak is related to the baseline f_{SOAE}. The frequency variation is mostly within a 40 Hz bandwidth between the 5th and 95th percentiles. Frequency bin=1 Hz. (D) A period modulation pattern obtained by averaging the f_{SOAE} waveform over multiple biasing cycles. The minima and maxima of the W-shaped pattern correspond to the zero crossings and extremes of the bias tone. Note that segments of the f_{SOAE} waveform with variability beyond the 90% interval are not included in the averaging. The ratio of included cycles vs the total number of biasing periods is indicated. τ is the averaged time delay of the f_{SOAE} modulation pattern.

tion in biasing pressure: a decrease in frequency excursion and prolonged delays of f_{SOAE} extremes (Fig. 10). As the L_{bias} was attenuated about 20 dB from 97 dB SPL, the range of f_{SOAE} oscillation subsided from about 25 to just 10 Hz. The frequency minima could be lower than the baseline value. The frequency increases at the biasing pressure extremes were not even. It was noticeable that the peaks and troughs of the f_{SOAE} were delayed with respect to the bias tone as the L_{bias} decreased. The averaged time delay (τ) of the f_{SOAE} maxima and minima was only 0.3 ms at 97 dB SPL L_{bias} and nearly 3.6 ms at the 77 dB SPL condition.

The period modulation pattern of a SOAE also varied with the f_{bias} (Fig. 11). As the f_{bias} increased from 25 to 100 Hz, the range of f_{SOAE} oscillation doubled from 30 to 60 Hz. Larger frequency oscillation was also marked by the minima below baseline value at zero crossings of the bias tone. Moreover, the typical FM patterns were observed at lower bias tone levels when the f_{bias} was higher. For example, the L_{bias} of 91 dB SPL at 25 Hz that induced a FM had to be attenuated by 10 dB when the f_{bias} was increased to 100 Hz. This was consistent with the two-tone interactions within the cochlea, i.e., suppression and distortion are more effective when the two frequencies are close. Thus, the bias tones with higher frequencies were more effective in producing f_{SOAE} oscillations. The effect of f_{bias} on delays of the f_{SOAE} pattern was variable. A slight reduction in the averaged time delay was shown as the f_{bias} increased. Moreover, the typical f_{SOAE} modulation patterns were more commonly found at lower biasing frequencies.

IV. DISCUSSION

A. Upward frequency shift of SOAE

The general effect of low-frequency biasing on f_{SOAE} is an upward shift. The analysis reveals that the frequency shift can reach up to about 2% of the baseline value depending on the bias tone power and the strength of a SOAE (Figs. 2–6). Consider the suppression of SOAE reported earlier (Bian and Watts, 2008), the biasing effects are opposite on SOAE magnitude and frequency, i.e., a reduction in SOAE amplitude is accompanied by an increase in frequency (Figs. 2–4). This is consistent with other observations using an external tone to suppress SOAEs (Long *et al.*, 1991, 1993; Norrix and Glattke, 1996). These studies show that when the suppressor frequency is below a SOAE, it can "push" the f_{SOAE} away from its original value, i.e., an upward shift. Such opposite changes in SOAE level and frequency seem to correlate with the effect of varying the inner ear pressure. Increasing and decreasing the ear-canal air pressure can change the f_{SOAE} (Schloth and Zwicker, 1983; Hauser *et al.*, 1993) similar to the quasistatic modulation patterns (Figs. 7 and 8). The canal pressure change in the magnitude of several kilopascals designed to modify the middle ear properties would unavoidably influence the inner ear pressure and the f_{SOAE}. Another indirect way of varying the inner ear pressure can be achieved by a posture change because the intracranial pressure could be transmitted to the inner ear fluid via the cochlear aqueduct. The f_{SOAE} can be shifted up to 2% with a reduction in amplitude when the body is tilted from upright to supine position (de Kleine *et al.*, 2000). Most of these

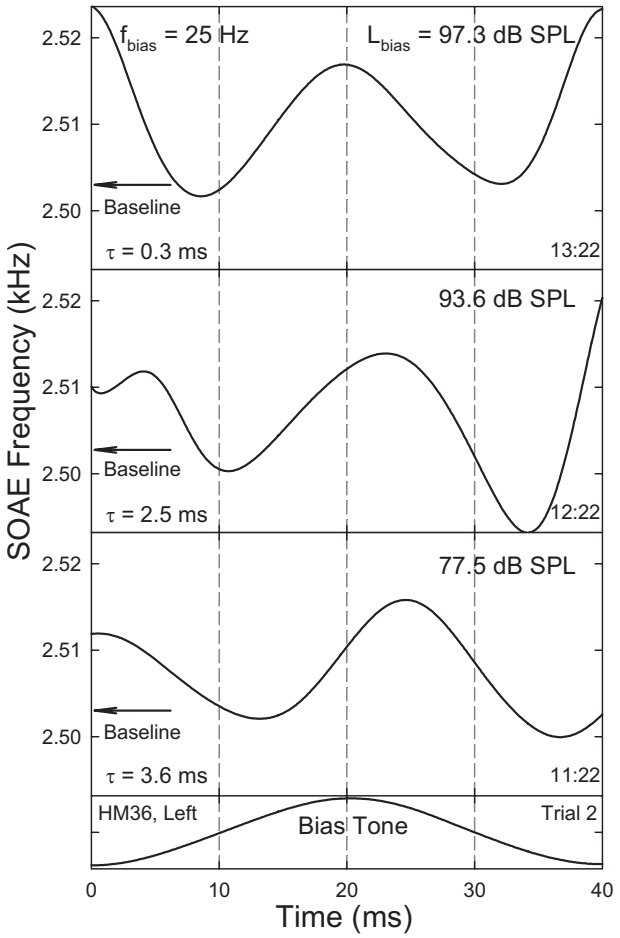


FIG. 10. Period modulation patterns: effects of L_{bias} on f_{SOAE}. From top to bottom, as L_{bias} decreases, the frequency excursion becomes smaller and the temporal features of the FM pattern lag the bias tone (bottom trace). The time delay (τ) averaged across the f_{SOAE} maxima and minima is indicated in each panel. For each half biasing cycle, the f_{SOAE} shows a minimum below the baseline value (arrow) following a bias tone zero crossing. The ratio of included cycles over the total biasing periods is indicated. The baseline f_{SOAE} is 2503 Hz.

frequency shifts are less than 20 Hz which is comparable to the present study (<50 Hz) since the biasing effects could be greater than a posture change. These observations indicate that the SOAE amplitude is largest and the frequency is lowest when the inner ear pressure is optimally balanced.

Pressure change in the inner ear could alter certain mechanical properties of the structures that are initializing or reflecting SOAE standing waves. It is generally agreed that the cochlear partition could be considered as a series of single-degree-of-freedom oscillators (Dallos, 1996). The natural frequency (ω_n) of each element is determined by the stiffness (k) and mass (m) or $\sqrt{k/m}$. It is apparent that to increase ω_n, one could either raise k or decrease m. Applying a bias tone could generally stiffen the cochlear partition, thus increasing the f_{SOAE}. If a group of oscillators at a particular section on the cochlear partition contribute to form a SOAE, then more apical elements in the group could be suppressed by the bias tone. This would result in a shift of contribution to relatively basal oscillators with a smaller net cellular m and larger k. The 1%–2% SOAE frequency shift (Fig. 6) seems to indicate a bandwidth that the group of oscillators is tuned to. If the bandwidth represents the boundary on the cochlear partition where OHCs are active components, then it is possible to estimate the population of the OHCs contributing to the SOAE. According to a cochlear frequency map (Greenwood, 1990), the characteristic frequency (f_{CF}) location as a distance (x) in millimeters from the apex can be calculated: $x=(1/a)\log_{10}(f_{CF}/A+k)$, where $a=0.06$ mm^{-1}, $A=165$ Hz, and $k=1$ for humans. Consider a frequency shift of 2%, the distance between two locations with a frequency ratio of 1.02 can be obtained:

$$\Delta x = \frac{1}{a}\log_{10}\left(\frac{1.02f_{CF}+A}{f_{CF}+A}\right). \quad (4)$$

This distance ranging from 0.1 mm at 500 Hz to about 0.14 mm for higher frequencies coincides with the extent of efferent innervations of a single fiber in the organ of Corti which rarely exceeds 150 μm (Robertson and Gummer, 1985). If the diameter of an OHC is 8 μm, then this range of distance on the cochlear partition could hold about 12–18 OHCs in one row or a total of 36–54 cells for three rows. It seems to suggest that irregularities in efferent connections to the OHCs may give rise to SOAEs. It also indicates that the SOAE could be present in a relatively restricted region in the cochlea. This localized SOAE production could be evident by the limited efficiency of the bias tone in influencing SOAEs with higher frequencies (Figs. 5 and 6). The 8 dB/octave increment of power demand for the bias tone to produce an effect on SOAEs (Fig. 6; Bian and Watts, 2008) and DPOAEs (Hensel *et al.*, 2007; Marquardt *et al.*, 2007; Bian and Scherrer, 2007) is inline with the tails of the mechanical tuning curves of the basilar membrane. Indeed, spontaneous oscillation has been expectedly observed at a specific location on the basilar membrane in a guinea pig with a SOAE (Nuttall *et al.*, 2004).

B. Opposite changes of SOAE magnitude and frequency: A transduction mechanism

The opposite biasing effects of SOAE amplitudes and frequencies are consistent with other observations in humans: correlated fluctuation of SOAE amplitude and frequency over time (Whitehead, 1991), contralateral acoustical stimulation (Mott *et al.*, 1989), and ipsilateral noise exposure (Kemp, 2007). Similar opposite behaviors of SOAE are found in manipulations of inner ear Ca^{2+} concentration in skinks (Manley *et al.*, 2004), noise exposure in lizards (Manley, 2006), dc injection in frogs (Wit *et al.*, 1989), and hypoxia in guinea pigs (Ohyama *et al.*, 1991). These evidences imply that the SOAE magnitude and frequency undergo correlated and opposite changes (van Dijk *et al.*, 1994) when the inner ear function is modified. Despite the drastic differences in inner ear anatomies among different species, a similar structural and functional element is the hair cell transduction which converts sound vibrations to electrical signals. In fact, there is an electrical counterpart of SOAE that can be recorded with an electrode placed in the inner ear (Wit *et al.*, 1989; Ohyama *et al.*, 1991). Therefore, the hair cell transduction could contribute to the generation of SOAEs.

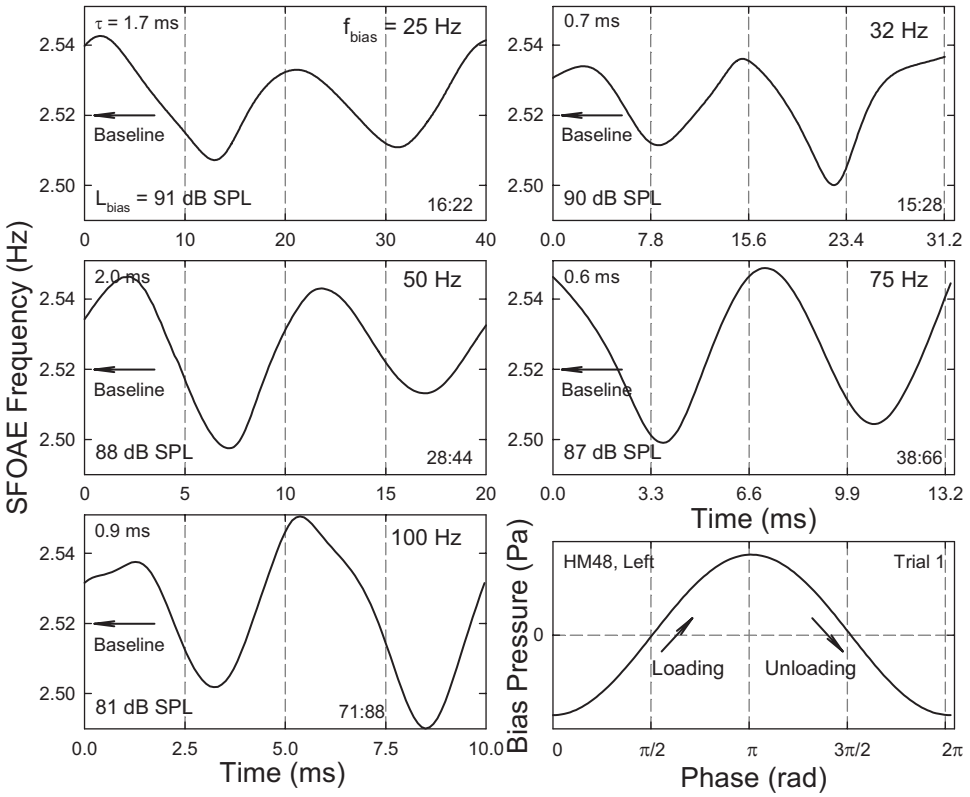


FIG. 11. Period modulation patterns: effects of f_{bias} on a SOAE. As f_{bias} increases, the frequency excursion becomes larger and the typical modulation patterns can be observed at lower biasing levels. Note that the f_{SOAE} minima can be lower than the baseline value (arrow). The time delay (τ) of the f_{SOAE} is slightly reduced with the f_{bias}. Numbers of included cycles are marked.

The suppression and modulation of SOAE amplitude can be interpreted by an operating point shift along the sigmoid-shaped cochlear transducer function which has a maximal gain at the inflection point (Bian and Watts, 2008). In the case of f_{SOAE} change, it could be attributed to the modulation of the stiffness in a section of the cochlear partition that is involved in the formation of a SOAE. The most variable constitute of the cochlear partition stiffness is probably contributed by the stereocilial bundles of OHCs because of their critical position between the tectorial membrane and reticular lamina where the mechanical deformation is maximal. Moreover, about 30%–40% of the hair bundle stiffness can be attributed to the gating springs (Kros, 1996), the molecular linkages that open and close the transduction channels. Therefore, the gating and adaptation motors in the hair bundles could be a factor responsible for the f_{SOAE} shift under low-frequency biasing.

The most important feature of the gating stiffness of hair bundles is the dip around their resting positions where 50% of transduction channels are open (Markin and Hudspeth, 1995). The relative hair bundle stiffness change (Δk) from unity can be determined by the probabilities of channel opening (P_0) and closing ($1-P_0$), where P_0 is dominated by a saturating nonlinearity, such as a Boltzmann function. If hair cells contribute to half of the cochlear partition stiffness, for constant m, the change in resonant frequency ($\Delta\omega$) is proportional to the square root of Δk, i.e.,

$$\Delta\omega\ (\%) \propto \sqrt{1-0.5\delta P_0(1-P_0)}, \tag{5}$$

where δ=0.3–0.4 represents the portion of the bundle stiffness contributed by the nonlinear gating spring. This relation qualitatively predicts the U-shaped quasistatic modulation pattern of f_{SOAE} with a $\Delta\omega$ of 1%–2% which is consistent with the experimental results. The shape of the gating stiffness depicts a second-order nonlinearity which was used in a van der Pol oscillator to model the fluctuations of SOAE amplitude and frequency (van Dijk and Wit, 1990). Increase in stiffness with displacement, a feature of the Duffing oscillator exhibiting a frequency shift (Talmadge *et al.*, 1991), can explain the FM of tone-evoked cochlear microphonic generated by the OHCs (Chertoff *et al.*, 2000). Therefore, both the amplitude and frequency of SOAEs could be related to the nonlinearity in the cochlear transducer.

C. Temporal effects

The dynamic aspect of the low-frequency biasing is represented by the temporal modulation of the f_{SOAE}. The FM typically occurs at relatively lower bias tone levels (75–95 dB SPL) when there is no significant AM, whereas typical AM is observed at higher levels (>95–100 dB SPL). This may indicate that the f_{SOAE} is more sensitive than the amplitude to mechanical alterations in the inner ear or less susceptible to noise contamination (Long and Talmadge, 1997). Tracking the SOAE amplitudes and frequencies over time with ipsi- or contralateral acoustic stimulations indicates that most of the f_{SOAE} reductions are statistically significant compared to nonsignificant SOAE amplitude increases (Whitehead, 1991). These changes are hypothesized to be a micromechanical effect of resetting the operating point of the hair cell transducer to its optimal position after an efferent activation (Bian, 2004). Such a high sensitivity of f_{SOAE} to minor alterations in cochlear mechanics, as first noticed by Zurek (1981) and later by Bell (1992), is also reflected by the observation that it can be modulated by heartbeat (Ren *et al.* 1995; Long and Talmadge, 1997). This may imply that subtle

alterations in mechanical properties of the cochlear transducer by operating point shifts could be a mechanism for cochlear gain control since mechanical structures with high gain may be required for SOAE formations.

Another temporal observation is the time delay in the f_{SOAE} oscillation with reference to the bias tone phase. The time delays increase with the reduction in the L_{bias} (Fig. 10). This effect is not typically observed in SOAE amplitudes because of a lack of AM at lower biasing levels. However, observations at these levels reveal a level-dependent reduction in frequency excursion or bandwidth. It can be noted that the FM time delay is inversely related to the range of f_{SOAE} oscillation. In a study of SOAE magnitude fluctuations, the relaxation or adaptation time constant is derived as the inverse of the bandwidth of the SOAE envelope which shows a low-pass characteristic (Wit, 1993). Similarly, the bandwidth of the f_{SOAE} oscillation can be determined by two factors: the rate of modulation that equals $2f_{bias}$ [Fig. 9(B)] and the Δf_{SOAE} that can be found from either the 90% range in the f_{SOAE} histogram [Fig. 9(C)] or twice the overall f_{SOAE} shift (Fig. 6, top). Thus, the delay time constant (τ) could be estimated:

$$\tau = \frac{1}{2\pi} \frac{1}{2\sqrt{\Delta f_{SOAE} f_{bias}}}. \tag{6}$$

Given a Δf_{SOAE} ranging from 20 to100 Hz, the delay constants vary from 3.5 to 0.8 ms from low to high biasing frequencies. These time constants are in close agreement with the observed values in Figs. 10 and 11, those for amplitudes of SOAEs (Bian and Watts, 2008), and DPOAEs (Bian *et al.*, 2004; Bian and Scherrer, 2007; Hensel *et al.*, 2007). The short-time delays are comparable with the onset delays of SOAE suppression (Schloth and Zwicker, 1983; Murphy *et al.*, 1995) and the auditory nerve responses (van der Heijden and Joris, 2005). These results may imply that a time-dependent mechanism, possibly the adaptation of hair cell transduction channels, could play a role in optimizing cochlear function and lead to the generation of SOAEs.

V. SUMMARY AND CONCLUSION

In addition to AM and suppression, low-frequency biasing of cochlear structures can shift and modulate the f_{SOAE}. The overall effect is an upward shift of the f_{SOAE} with increase in the L_{bias}. Up to 2% frequency shifts are observed depending on the relative strength of a bias tone over a particular SOAE based on the biasing frequency and level. The frequency shift usually occurs at relatively lower bias tone levels prior to the AM and suppression of the SOAE. Analyzing the f_{SOAE} at the bias tone peaks and troughs reveals a U-shaped modulation pattern which indicates that biasing in either positive or negative direction can increase the f_{SOAE}. The instantaneous f_{SOAE} over one biasing cycle demonstrates a W shape. The f_{SOAE} peaks at the biasing extremes and minimizes near the zero crossings of the bias tone with a short delay. These static and dynamic effects of the bias tone on the f_{SOAE} indicate that modifications of the mechanical properties of the hair cell transducer could likely cause the frequency shift and modulation. Periodic variation in possibly the stiffness of OHC hair bundles due to biasing may imply that the regulation of cochlear transducer feedback gain is achieved by cellular actions, such as adaptations. The results suggest that the nonlinearity in cochlear transducers acting as internal reflection interfaces of an inner ear standing wave (Shera, 2003) could influence the ear-canal SOAE properties, such as amplitude stability and spread of the spectral peak width.

ACKNOWLEDGMENTS

Kelly Watts and Nicole Scherrer are acknowledged for assisting in recruiting the subjects and in data collection. Mark Chertoff, Shixiong Chen, Jeff Lichtenhan, and Chris Bergevin provided many discussions on various issues related to this project. The clinical staff in the Department of Speech and Hearing Science at ASU is appreciated for sharing their equipment. This work was supported by Grant No. R03 DC006165 from the National Institute on Deafness and Other Communication Disorders of the NIH.

APPENDIX

If a carrier signal with a frequency of f_c is frequency modulated by a low-frequency tone, the FM signal can be written as

$$s_{FM}(t) = A \sin(2\pi f_c t + I_m \sin 2\pi f_m t), \tag{A1}$$

where A is the amplitude of the carrier, f_m is the modulator frequency ($f_m \ll f_c$), and I_m is the modulation index which is defined as the ratio of the frequency change or deviation (Δf) with respect to the modulation frequency, i.e.,

$$I_m = \Delta f / f_m. \tag{A2}$$

Applying trigonometric identity $\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$, and replacing frequencies $2\pi f_c$ and $2\pi f_m$ with their angular forms ω_c and ω_m, the FM signal in Eq. (A1) can be rewritten as

$$s_{FM}(t) = A \sin\omega_c t \cos(I_m \sin\omega_m t) + A \cos\omega_c t \sin(I_m \sin\omega_m t). \tag{A3}$$

According to Bessel function identities: $\cos(x\sin\theta) = 2\Sigma_{n=0}^{\infty} J_n(x)\cos(n\theta)$, where n is even, and $\sin(x\sin\theta) = 2\Sigma_{n=1}^{\infty}(-1)^n J_n(x)\sin(n\theta)$, where n is odd, the spectral components of the FM signal around the carrier can be expressed as follows:

$$S_{FM}(f) = A \sin\omega_c t \left[2\sum_{n=0}^{\infty} J_n(I_m)\cos(n\omega_m t) \right] + A \cos\omega_c t \left[2\sum_{n=1}^{\infty} (-1)^n J_n(I_m)\sin(n\omega_m t) \right]. \tag{A4}$$

Multiplying through, expanding the first term as $n=0$, and applying the formula $\sin\alpha\cos\beta = \frac{1}{2}\sin(\alpha\pm\beta)$ for the rest of the terms, it yields the spectral expression of the carrier and sideband components

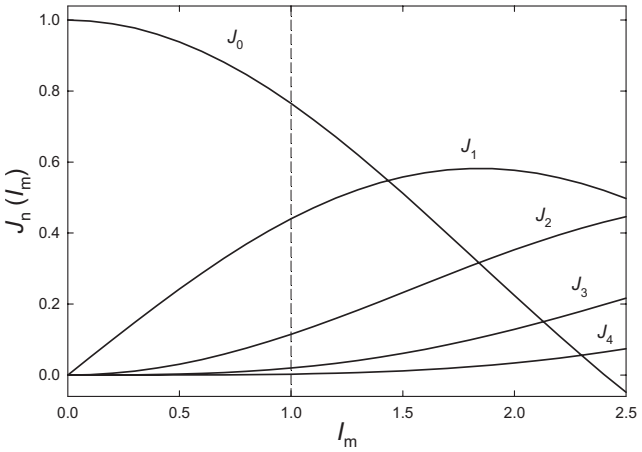


FIG. 12. Sideband amplitudes of FM signals. As the modulation index (I_m) increases to 2, the sideband amplitudes ($J_1, J_2, \ldots$) also increase. When $I_m < 1$, the amplitudes of sidebands with orders higher than 2 are negligible (<2%). Calculation is based on summing the first 15 terms of the Bessel function of the first kind in Eq. (A7). J_0 is the amplitude of the carrier, J_1 is the amplitude of sideband I, J_2 is the amplitude of sideband II, and etc.

$$S_{\mathrm{FM}}(f) = J_0(I_m)\sin\,\omega_c t + \sum_{n=2}^{\infty} J_n(I_m)\sin(\omega_c \pm n\omega_m t) - \sum_{n=1}^{\infty} (-1)^n J_n(I_m)\sin(\omega_c \pm n\omega_m t). \quad \text{(A5)}$$

Combining the odd and the even terms, the above equation can be written as

$$S_{\mathrm{FM}}(f) = J_0(I_m)\sin\,\omega_c t + \sum_{n=1}^{\infty} J_n(I_m)\sin(\omega_c + n\omega_m)t + (-1)^n \sum_{n=1}^{\infty} J_n(I_m)\sin(\omega_c - n\omega_m)t. \quad \text{(A6)}$$

The first term is the frequency component of the carrier with an amplitude of J_0. The second and third terms contain all the components in the USBs and LSBs of the FM signal. The sidebands are distributed at integer (n) multiples of the f_m above and below the carrier. The amplitudes of the carrier and the sidebands are determined by the coefficient $J_n(I_m)$ in the form of the Bessel function of the first kind:

$$J_n(I_m) = \sum_{i=0}^{\infty} (-1)^i \frac{(I_m/2)^{2i+n}}{i\,!\,\cdot\,\Gamma(i+n+1)}, \quad \text{(A7)}$$

where n denotes the order of the sidebands, with 0 indicating the carrier component, and Γ is a gamma function which for all positive integers is equivalent to a factorial, i.e., $\Gamma(i+n+1)=(i+n)!$. For varying I_m, the Bessel function is solved by summing 15 terms in Eq. (A7) and plotted in Fig. 12. It can be observed that when I_m is smaller than 1 which is the case of low-frequency modulation of SOAEs, the sizes of sidebands with orders higher than 2 (J_3 and J_4) are negligible and can be omitted in the spectral windowing. Therefore, the bandwidth of the FM signal can be determined as

$$B_m = 2(\Delta f + 2f_m), \quad \text{(A8)}$$

where Δf is added to account for the shift of f_c due to the low-frequency biasing.

Ashmore, J. (**2008**). "Cochlear outer hair cell motility," Physiol. Rev. **88**, 173–210.

Bargones, J. Y., and Burns, E. M. (**1988**). "Suppression tuning curves for spontaneous otoacoustic emissions in infants and adults," J. Acoust. Soc. Am. **83**, 1809–1816.

Bell, A. (**1992**). "Circadian and menstrual rhythms in frequency variations of spontaneous otoacoustic emissions from human ears," Hear. Res. **58**, 91–100.

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., and Chertoff, M. E. (**2006**). "Modulation patterns and hysteresis: Probing cochlear dynamics with a bias tone," in *Auditory Mechanisms: Processes and Models*, edited by A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer (World Scientific, Singapore), pp. 93–100.

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., Linhardt, E. E., and Chertoff, M. E. (**2004**). "Cochlear hysteresis: Observation with low-frequency modulated distortion product otoacoustic emissions," J. Acoust. Soc. Am. **115**, 2159–2172.

Bian, L., and Scherrer, N. M. (**2007**). "Low-frequency modulation of distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **122**, 1681–1692.

Bian, L., and Watts, K. L. (**2008**). "Effects of low-frequency biasing on spontaneous otoacoustic emissions: Amplitude modulation," J. Acoust. Soc. Am. **123**, 887–898.

Bilger, R. C., Matthies, M. L., Hammel, D. R., and Demorest, M. E. (**1990**). "Genetic-implications of gender differences in the prevalence of spontaneous otoacoustic emissions," J. Speech Hear. Res. **33**, 418–432.

Chertoff, M. E., Lerner, D., Amani-Taleshi, D., and Nagai, Y. (**2000**). "Characterizing non-linearity in the cochlear microphonic using the instantaneous frequency," Hear. Res. **145**, 190–202.

Dallos, P. (**1996**). "Overview: Cochlear neurobiology," in *The Cochlea*, edited by P. Dallos, A. N. Propper, and R. R. Fay (Springer, New York), pp. 1–43.

de Kleine, E., Wit, H. P., van Dijk, P., and Avan, P. (**2000**). "The behavior of spontaneous otoacoustic emissions during and after postural changes," J. Acoust. Soc. Am. **107**, 3308–3316.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species-29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hauser, R., Probst, R., and Harris, F. P. (**1993**). "Effects of atmospheric pressure variation on spontaneous, transiently evoked, and distortion product otoacoustic emissions in normal human ears," Hear. Res. **69**, 133–145.

Hensel, J., Scholz, G., Hurttig, U., Mrowinski, D., and Janssen, T. (**2007**). "Impact of infrasound on the human cochlea," Hear. Res. **233**, 67–76.

Kemp, D. T. (**2007**). "The basics, the science, and the future potential of otoacoustic," in *Otoacoustic Emissions: Clinical Applications*, 3rd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 7–42.

Khalfa, S., and Collet, L. (**1996**). "Functional asymmetry of medial olivocochlear system in humans. Towards a peripheral auditory lateralization," NeuroReport **7**, 993–996.

Kros, C. J. (**1996**). "Physiology of mammalian cochlear hair cells," in *The Cochlea*, edited by P. Dallos, A. N. Propper, and R. R. Fay (Springer, New York), pp. 318–385.

Kuroda, T. (**2007**). "Clinical investigation on spontaneous otoacoustic emissions (SOAE) in 447 ears," Auris Nasus Larynx **34**, 29–38.

Long, G. R., Sun, C., and Talmadge, C. L. (**1993**). "Interaction between spontaneous otoacoustic emissions and external tones: Supression, frequency shifting and distortion product generation," 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. 40–46.

Long, G. R., and Talmadge, C. L. (**1997**). "Spontaneous otoacoustic emission frequency is modulated by heartbeat," J. Acoust. Soc. Am. **102**, 2831–2848.

Long, G. R., Tubis, A., and Jones, K. L. (**1991**). "Modeling synchronization

and suppression of spontaneous otoacoustic emissions using Van der Pol oscillators: Effects of aspirin administration," J. Acoust. Soc. Am. **89**, 1201–1212.
Manley, G. A. (**2006**). "Spontaneous otoacoustic emissions from free-standing stereovillar bundles of ten species of lizard with small papillae," Hear. Res. **212**, 33–47.
Manley, G. A., Sienknecht, U., and Köppl, C. (**2004**). "Calcium modulates the frequency and amplitude of spontaneous otoacoustic emissions in the bobtail skink," J. Neurophysiol. **92**, 2685–2693.
Markin, V. S., and Hudspeth, A. J. (**1995**). "Gating-spring models of mechanoelectrical transduction by hair cells of the internal ear," Annu. Rev. Biophys. Biomol. Struct. **24**, 59–83.
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.
Martin, P., Bozovic, D., Choe, Y., and Hudspeth, A. J. (**2003**). "Spontaneous oscillation by hair bundles of the bullfrog's sacculus," J. Neurosci. **23**, 4533–4548.
McFadden, D. (**2002**). "Masculinization effects in the auditory system," Arch. Sex Behav. **31**, 99–111.
Moulin, A. (**2000**). "Influence of primary frequencies ratio on distortion product otoacoustic emissions amplitude, II: Interrelations between multicomponent DPOAEs, tone-burst-evoked OAEs, and spontaneous OAEs," J. Acoust. Soc. Am. **107**, 1471–1486.
Mott, J. B., Norton, S. J., Neely, S. T., and Warr, W. B. (**1989**). "Changes in spontaneous otoacoustic emissions produced by acoustic stimulation of the contralateral ear," Hear. Res. **38**, 229–242.
Murphy, W. J., Tubis, A., Talmadge, C. L., and Long, G. R. (**1995**). "Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones, II: Suppression of interacting emissions," J. Acoust. Soc. Am. **97**, 3711–3720.
Norrix, L. W., and Glattke, T. J. (**1996**). "Distortion product otoacoustic emissions created through the interaction of spontaneous otoacoustic emissions and externally generated tones," J. Acoust. Soc. Am. **100**, 945–955.
Nuttall, A. L., Grosh, K., Zheng, J., de Boer, E., Zou, Y., and Ren, T. (**2004**). "Spontaneous basilar membrane oscillation and otoacoustic emission at 15 kHz in a guinea pig," J. Assoc. Res. Otolaryngol. **5**, 337–348.
O'Brien, A. J. (**1994**). "Temperature dependency of the frequency and level of a spontaneous otoacoustic emission during fever," Br. J. Audiol. **28**, 281–290.
Ohyama, K., Wada, H., Kobayashi, T., and Takasaka, T. (**1991**). "Spontaneous otoacoustic emissions in the guinea pig," Hear. Res. **56**, 111–121.
Ozturan, O., and Oysu, C. (**1999**). "Influence of spontaneous otoacoustic emissions on distortion product otoacoustic emission amplitudes," Hear. Res. **127**, 129–136.
Pasanen, E. G., and McFadden, D. (**2000**). "An automated procedure for identifying spontaneous otoacoustic emissions," J. Acoust. Soc. Am. **108**, 1105–1116.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.
Pujol, R. (**2001**). "Neural anatomy of the cochlea: Development and plasticity," in *Physiology of the Ear*, 2nd ed., edited by A. F. Jahn and J. Santos-Sacchi (Singular, New York), pp. 515–528.
Rabinowitz, W. M., and Widin, G. P. (**1984**). "Interaction of spontaneous otoacoustic emissions and external sounds," J. Acoust. Soc. Am. **76**, 1713–1720.
Ren, T., Zhang, M., Nuttall, A. L., and Miller, J. M. (**1995**). "Heart beat modulation of spontaneous otoacoustic emissions in guinea pig," Acta Oto-Laryngol. **115**, 725–731.
Robertson, D., and Gummer, M. (**1985**). "Physiological and morphological characterization of efferent neurones in the guinea pig cochlea," Hear. Res. **20**, 63–77.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Schloth, E., and Zwicker, E. (**1983**). "Mechanical and acoustical influences on spontaneous oto-acoustic emissions," Hear. Res. **11**, 285–193.
Shera, C. A. (**2003**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized occhlear standing waves," J. Acoust. Soc. Am. **114**, 244–262.
Talmadge, C. L., Long, G. R., Murphy, W. J., and Tubis, A. (**1993**). "New off-line method for detecting spontaneous otoacoustic emissions in human subjects," Hear. Res. **71**, 170–182.
Talmadge, C. L., Tubis, A., Wit, H., P., and Long, G., R. (**1991**). "Are spontaneous otoacoustic emissions generated by self-sustained cochlear oscillators?," J. Acoust. Soc. Am. **89**, 2391–2399.
Thiers, F. A., Burgess, B., J., and Nadol, Jr., J. B (**2002**). "Reciprocal innervation of outer hair cells in a human infant," J. Assoc. Res. Otolaryngol. **3**, 269–278.
van der Heijden, M., and Joris, P. X. (**2005**). "The speed of auditory low-side suppression," J. Neurophysiol. **93**, 201–209.
van Dijk, P., and Wit, H. P. (**1990**). "Amplitude and frequency fluctuations of spontaneous otoacoustic emissions," J. Acoust. Soc. Am. **88**, 1779–1793.
van Dijk, P., Wit, H. P., Tubis, A., Talmadge, C. L., and Long, G. R. (**1994**). "Correlation between amplitude and frequency fluctuations of spontaneous otoacoustic emissions," J. Acoust. Soc. Am. **96**, 163–169.
Whitehead, M. L. (**1991**). "Slow variations of the amplitude and frequency of spontaneous otoacoustic emissions," Hear. Res. **53**, 269–280.
Wit, H. P. (**1993**). "Amplitude fluctuations of spontaneous otoacoustic emissions caused by internal and externally applied noise sources," Prog. Brain Res. **97**, 59–65.
Wit, H. P., van Dijk, P., and Segenhout, J. M. (**1989**). "DC injection alters spontaneous otoacoustic emission frequency in the frog," Hear. Res. **41**, 199–204.
Zurek, P. M. (**1981**). "Spontaneous narrowband acoustic signals emitted by human ears," J. Acoust. Soc. Am. **69**, 514–523.

Auditory brainstem responses to a chirp stimulus designed from derived-band latencies in normal-hearing subjects

Claus Elberling[a)]
Oticon A/S, "Eriksholm," DK-3070 Snekkersten, Denmark

Manuel Don
Electrophysiology Laboratory, House Ear Institute, Los Angeles, California 90057

(Received 1 July 2008; revised 27 August 2008; accepted 27 August 2008)

In an attempt to compensate for the temporal dispersion in the human cochlea, a chirp has previously been designed from estimates of the cochlear delay based on derived-band auditory brain-stem response (ABR) latencies [Elberling *et al.* (2007). "Auditory steady-state responses to chirp stimuli based on cochlear traveling wave delay," J. Acoust. Soc. Am. **122**, 2772–2785]. To evaluate intersubject variability and level effects of such delay estimates, a large dataset is analyzed from 81 normal-hearing adults (fixed click level) and from a subset thereof (different click levels). At a fixed click level, the latency difference between 5700 and 710 Hz ranges from about 2.0 to 5.0 ms, but over a range of 60 dB, the mean relative delay is almost constant. Modeling experiments demonstrate that the derived-band latencies depend on the cochlear filter buildup time and on the unit response waveform. Because these quantities are partly unknown, the relationship between the derived-band latencies and the basilar membrane group delay cannot be specified. A chirp based on the above delay estimates is used to record ABRs in ten normal-hearing adults (20 ears). For levels below 60 dB nHL, the gain in amplitude of chirp-ABRs to click-ABRs approaches 2, and the effectiveness of chirp-ABRs compares favorably to Stacked-ABRs obtained under similar conditions. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2990709]

PACS number(s): 43.64.Qh, 43.64.Bt, 43.64.Ri [BLM] Pages: 3022–3037

I. INTRODUCTION

A. The chirp stimulus

The concept of the chirp was first applied to auditory electrophysiology by Shore and Nutall (1985), and has since been studied intensively for its use within the auditory field; for a recent review see, e.g., Elberling *et al.* (2007). A chirp stimulus—or more specifically an upward chirp—is designed to compensate for the temporal dispersion in the cochlea related to the traveling wave delay, e.g., Shore and Nutall (1985), Lütkenhöner *et al.* (1990), and Dau *et al.* (2000). In response to a brief stimulus, as, for example, a click, the cochlear traveling wave takes some time to reach from the base of the cochlea to its apical end. Therefore, the different neural units along the cochlear partition will not be stimulated at the same time and the neural activity across all nerve fibers will be smeared. This lack of temporal synchrony can be partly neutralized by an upward chirp stimulus, in which higher frequency components are delayed relative to lower frequency components. The design of such a chirp must be based on a model of the cochlear delay as it is reflected in the broadband click response—being either the auditory compound action potential (ACAP) or the auditory brain-stem response (ABR). It has been shown repeatedly that a chirp is more efficient than a corresponding click both in the recording of the ACAP, the ABR, and of the auditory steady-state response (ASSR)—see, e.g., Elberling *et al.* (2007).

[a)]Present address: Oticon A/S, "Eriksholm," 243 Kongevejen, DK-3070 Snekkersten, Denmark. Electronic mail: ce@oticon.dk

B. Latency-frequency functions obtained from derived-band ABRs

One of the chirps used in the ASSR-study by Elberling *et al.* (2007) was based on latencies from derived-band ABRs originally obtained from a group of normal-hearing subjects by Don *et al.* (2005). The mean latencies were fitted to a power-function model, which described the average latency-frequency function or the cochlear delay in normal-hearing adults. More specifically, the latency model was used to represent the group delay of a linear cochlear two-port system that modifies the phase of the applied stimulus—e.g., a click. Subsequently, the inverse group delay was used to design a chirp that theoretically should be able to compensate for the delay introduced by the cochlea.

The individual derived-band ABR latencies—described above—showed a relatively large variation across the group of normal-hearing individuals, and further it was unknown whether the power-function model (originally described by Anderson *et al.*, 1971) was adequate in describing the individual latency-frequency relationship.

This raises therefore the following questions:

(1) What is the interindividual variation of the derived-band ABR latencies in a large group of normal-hearing adults?
(2) What are the implications of this variance?
(3) Is the power-function model a relevant descriptor of the latency-frequency function based on individual derived-band latencies?

 © 2008 Acoustical Society of America

C. Chirps based on derived-band ABR latencies at different stimulus levels

Fobel and Dau (2004) described an interesting effect of stimulus level in an ABR study on normal-hearing subjects using three different chirp stimuli. The results implied that two of the chirps designed using different values of the cochlear delay would be optimal at different ranges of stimulus levels. One of the chirps, the M-chirp, was based on cochlear delay values from a model of the mechanical characteristics of the cochlea (de Boer, 1980). Another chirp, the A-chirp, was based on cochlear delay values derived from the latencies of frequency-specific tone-burst ABRs in normal-hearing subjects (Neely *et al.*, 1988), which resulted in cochlear delays that changed significantly with stimulus level.

In the study by Elberling *et al.* (2007) a set of three chirps was used to record the ASSR at 50 and 30 dB nHL in 49 normal-hearing adults. The three chirps were as follows: (1) the chirp based on cochlear delay values in accordance with the above mentioned power-function model fitted to the mean derived-band ABR latencies obtained by Don *et al.* (2005). (2) The chirp based on the above mentioned tone-burst ABR latencies described by Neely *et al.* (1988), but only using the values for the cochlear delay that was described for one specific stimulus level. This level was approximately 65 dB nHL and the chirp based on Neely *et al.* (1988) turned out to correspond to the A-chirp at 50 dB SL used by Fobel and Dau (2004) (for further details see the discussion in Elberling *et al.*, 2007 p. 2781). (3) The chirp based on de Boer (1980) similar to the M-chirp used by Fobel and Dau (2004). In this study (Elberling *et al.*, 2007), it was found that the ASSRs in response to the chirp based on Don *et al.* (2005) were larger than the ASSRs to any of the other two chirps; however, this finding was only statistical significant at 50 dB nHL.

These results lead to the following question: What is the effect of stimulus level on the latency-frequency function described by the derived-band ABRs?

D. Derived-band ABR latencies and the traveling wave delay

In several publications about chirp stimuli, the latency-frequency functions that are used to design the chirps are thought to represent the cochlear traveling wave delay in normal-hearing adults (e.g., Dau *et al.*, 2000, Fobel and Dau, 2004, Stürzebecher *et al.*, 2006, Elberling *et al.*, 2007). Recently, however, Ruggero and Temchin (2007) analyzed the different latency-frequency functions obtained from electrophysiological recordings in humans—specifically tone-burst ABRs and derived-band ABRs—and questioned whether or not these functions were in accordance with their own estimate of the group delay of the basilar membrane (BM) traveling wave. For the derived-band responses, Ruggero and Temchin (2007) concluded that the precise relationship between BM delays and derived-band estimates cannot be specified at present with any certainty. With updated information about the latency-frequency function obtained from derived-band ABR-latencies in a large group of normal-hearing adults, the question should be revisited: How does such an average latency-frequency function relate to an estimate of the group delay of the BM traveling wave in humans?

E. The efficiency of a chirp based on derived-band ABR latencies

In the study by Elberling *et al.* (2007) the efficiency of a chirp based on derived-band ABR-latencies [chirp based on Don *et al.* (2005)] was evaluated by recording the ASSR (stimulus rate 90/s) from 49 normal-hearing adults. The chirp ASSRs were compared with the corresponding click ASSRs at two stimulus levels, viz 50 and 30 dB nHL. From the grand average ASSR waveforms obtained from the different recording conditions (see Elberling *et al.*, 2007, Fig. 6) the average gain (ratio) in peak-peak response amplitude between the chirp and the click can be evaluated: at 50 dB nHL the ratio was $820/340 \sim 2.4$, and at 30 dB nHL the ratio was $560/280 \sim 2.0$. These gain values are of the same order of magnitude as those found for the ABR by Fobel and Dau (2004, Fig. 4) using chirps designed in a different way.

In the paper by Elberling *et al.* (2007), the terms *input compensation* and *output compensation* were introduced. By input compensation, the temporal dispersion in the cochlea is compensated for by using a chirp stimulus, whereas by output compensation the cochlear delay is compensated for by means of the Stacked ABR technique. This latter technique is based on the derived-band method (Teas *et al.*, 1962) using the same underlying concept as the chirp, namely, to increase the temporal synchronization of activity across the different neural elements in the cochlea (for further description of the Stacked ABR technique see, e.g., Don *et al.*, 1994, 2009).

In the paper by Don *et al.* (2009), ABR recordings obtained previously from a group of normal-hearing subjects (Don *et al.*, 2005) were evaluated. Both click ABRs and derived-band ABRs were obtained at 60 dB nHL and three different Stacked ABRs were constructed from the individual recordings and compared to the corresponding click ABR. The different Stacked ABRs were formed by aligning the derived-band ABRs according to (1) the individual's peak latencies (named standard Stacked ABR), (2) the group mean latencies (named mean Stacked ABR), and (3) the modeled latencies in accordance with a power-function model fitted to the group mean latencies (named modeled Stacked ABR). This latter power-function model corresponds to the model used by Elberling *et al.* (2007) to design the chirp based on Don *et al.* (2005). The average gain (ratio) in peak-peak response amplitude between each of the three Stacked ABRs and the click was evaluated as follows: (1) for the standard Stacked ABR, the ratio was ~2.59; (2) for the mean stacked ABR, the ratio was ~2.36; and (3) for the model Stacked ABR, the ratio was ~2.27.

This leads to the following questions:

(1) What is the gain in peak-peak amplitude of the chirp ABR over the click ABR in normal-hearing adults?
(2) How do these gain values compare to those obtained by the Stacked ABR?
(3) What are the differences between the waveforms of the

click ABR, the chirp ABR, and the Stacked ABR obtained under comparable recording conditions in normal-hearing adults?

F. The aims of the present study

In order to address the questions raised above, there are four specific aims of the present study.

First, to create a combined dataset with derived-band ABR latencies obtained from a total of 81 normal-hearing subjects. All data are retrieved from previous experiments (Don *et al.*, 1998, 2005). This combined dataset will be used to describe the interindividual variations and to evaluate the effectiveness of using a power-function model to describe the latency-frequency relationship in the individual. Subsequently, the dataset will be used to design a chirp stimulus.

Second, to describe how the latency-frequency relationship varies with stimulus level over a range of 60 dB. The data underlying this analysis also originate from a previous experiment (Don *et al.*, 1998).

Third, to analyze the differences between the latencies of the derived-band responses and the estimated delays of the BM traveling wave. This analysis will include a modeling of the derived-band responses.

Fourth, to evaluate the efficiency of chirp ABRs (input compensation) with Stacked ABRs (output compensation). This will be achieved by evaluating the amplitude ratios between chirp and click ABRs recorded in normal-hearing subjects. The distributions of these ratios will be compared with similar distributions between Stacked and click ABRs that recently have been described by Don *et al.* (2009). Finally, the waveform morphology of click ABRs, chirp ABRs, and Stacked ABRs will be compared.

II. COMBINED DERIVED-BAND ABR LATENCY DATA

A. Two different datasets

Two datasets of derived-band ABR latencies are combined for the present evaluation. Dataset I consists of derived-band latencies recorded by Don *et al.* (1998) and dataset II consists of corresponding data subsequently recorded in the same laboratory at House Ear Institute by Don *et al.* (2005). Dataset II was used previously for the design of a chirp stimulus as reported by Elberling *et al.* (2007). Details about the recording procedures, etc., can be found in the original publications but will be briefly summarized in the following.

1. Subjects

Both investigations (Don *et al.*, 1998, 2005) received approval from the Institute Review Board (IRB). Prior to testing, the purpose and procedure of the study were orally presented to each test subject by the experimenter. The experimenter then answered any questions the subject had regarding his/her participation. Finally, each subject read and signed an IRB approved informed consent form.

The test group for dataset I (Don *et al.*, 1998) consisted of 42 normal-hearing subjects[1] (23 females and 19 males) with ages 18–38 years. The test group for dataset II (Don *et al.*, 2005) consisted of 39 normal-hearing subjects (20 females and 19 males) with ages 18–39 years. Normal hearing was defined as pure-tone thresholds of 10 dB HL or better for frequencies between 500 and 4000 Hz and 15 dB HL or better for 6000 and 8000 Hz.

2. Stimuli

The stimuli were rarefaction clicks produced by applying 100 μs electric pulses to the earphones. For the recording of dataset I, a TDH 50P earphone was used, whereas dataset II was obtained using an ER-2 insert earphone. The clicks were presented at a rate of about 45/s (interstimulus interval=22 ms) and at a level of approximately 60 dB nHL.[2]

Ipsilateral pink-noise masking was used to obtain the derived-band ABRs by means of the high-pass masking technique. The noise was presented at a level which was sufficient to mask the ABR to the clicks. There were six stimulus conditions: clicks presented alone (unmasked condition), and clicks presented with ipsilateral pink noise high-pass filtered at 8000, 4000, 2000, 1000, and 500 Hz. The slope of the high-pass filtered masking noise was 96 dB/octave.

3. ABR recordings

The subjects were placed in a reclining chair in a double-walled sound-treated room (IAC). ABRs were recorded differentially between electrodes applied to the vertex (C_z) and the ipsilateral mastoid (M_1 or M_2); the electrode at the contralateral mastoid was used as ground. The EEG was bandpass filtered from 100 to 3000 Hz using filter slopes of 12 dB/octave. ABRs were obtained using noise estimation techniques (Elberling and Don, 1984) and weighted averaging techniques (Elberling and Wahlgreen, 1985; Don and Elberling, 1994). These techniques reduce the destructive effects of episodic physiological background noise variation on the ABR average by weighting the average toward those blocks of sweeps with low estimated background noise. Data collection for a run was terminated when the estimated residual background noise in the average reached 20 nV (rms) or less. Thus, all recordings had approximately the same low residual background noise levels.

The derived-band responses were subsequently produced from ABR recordings first to a broadband click presented alone and then to a series of simultaneous ipsilateral presentations of the click and high-pass noise with varying cutoff frequencies as described above. By subtracting the response for one run from the previous, a derived-band response is formed. This method for generating a series of five derived-band ABRs has been presented before in a number of studies (Don and Eggermont, 1978; Parker and Thornton, 1978; Don *et al.*, 1997; 2005) and is based on the high-pass masking technique (Teas *et al.*, 1962). Each derived-band response represents octave-wide activity with a theoretical center frequency equal to the geometrical mean of the high-pass cutoff frequencies of the two maskers used to form the derived-band ABR.

TABLE I. Observed derived-band latencies (mean and standard deviation) from dataset I and II, as well as the calculated values for the combined dataset (N=81 Ss). The observed latencies are corrected for the wave I-V delay (4.10 ms) and the values in dataset II also for the acoustical delay (0.86 ms), in the ER-2 earphone. The column to the right shows the mean latencies of the combined dataset subtracted a value of 1.0 ms (for explanation see text).

Band center frequency (Hz)	1998 (N=42) latency (ms)		2005 (N=39) latency (ms)		Combined (N=81) latency (ms)		Corrected (N=81) latency (ms)
	Mean	SD	Mean	SD	Mean	SD	Mean-1.0 ms
Unmasked	2.13	0.27	2.15	0.27	2.14	0.27	
11300	2.09	0.36	2.10	0.27	2.09	0.32	
5700	2.14	0.29	2.19	0.35	2.17	0.32	1.17
2800	2.89	0.42	2.82	0.39	2.86	0.40	1.86
1400	4.07	0.58	3.79	0.54	3.93	0.56	2.93
710	5.66	0.73	5.48	0.85	5.57	0.79	4.57

4. Derived-band latencies

For both datasets, the wave V latencies of the unmasked click ABR and of the five derived-band responses were collected. In order to relate this activity to the level of the cochlea—e.g., the traveling wave and the action potentials in the peripheral part of the auditory nerve—the wave I-V latency difference of 4.1 ms (Elberling and Parbo, 1987) was subtracted from all observed latency values. Additionally, 0.86 ms was subtracted from the values in dataset II in order to compensate for the acoustical delay in the long tube of the ER-2 earphone (i.e., 292 mm). After these corrections were made, the Kolmogorov–Smirnov test of normality (Siegel, 1956) was performed on the values from each dataset and each condition. The results indicated that none of the data distributions deviates significantly from a Gaussian distribution described by the observed mean and standard deviation. Thus, it was assumed that all the data reported here are normally distributed, justifying the use of normal parametric statistics in the analyses. Mean and standard deviation for both studies are calculated and shown in Table I.

The two datasets produce mean latencies that differ very little from each other except for the 1400 Hz band, where the derived-band latency difference is 0.28 ms (4.07–3.79 ms) longer for dataset I; this difference is significant ($p<0.05$; t-test). Analysis of the frequency characteristics of the two earphones (amplitude and phase) reveals no obvious differences between their group delays in the 1000–2000 Hz frequency range, which could explain a latency difference of the observed magnitude (see Sec. IV A and endnote 3).[3] In the four derived bands: 5700, 2800, 1400, and 710 Hz, the variances (standard deviations) of the two datasets have the same order of magnitude and do not differ significantly from each other (NS, F-test). Since the observed difference for the 1400 Hz band is small (<0.30 ms), the two datasets can meaningfully be combined by calculating the weighted means and standard deviations of the two sets of latencies. These combined values are shown in Table I.

B. Describing the combined dataset

In the further analysis of the combined data, the values corresponding to the highest band center frequency, i.e., 11 300 Hz, are excluded; the values for this derived band are regarded to be uncertain because the level of activation is much lower in this frequency region given the ear's sensitivity curve and the falling spectrum of the click above 8000–9000 Hz (Don *et al.*, 1979; Elberling *et al.*, 2007). The remaining four data points represent an observed latency-frequency function.

Such latency functions of electrophysiological data can be described by a power function as suggested by Anderson *et al.* (1971):

$$\tau = k \cdot f^{-d}, \tag{1}$$

where τ is the latency in seconds, f is the frequency in Hertz, and k and d are constants. This formula was also used by Eggermont (1979) to describe derived-band ACAP latencies and was used recently by Elberling *et al.* (2007) to describe the latency-frequency functions of different sets of compound latency data.

The latency-frequency data for each individual test subject (N=81) as well as the mean values (Table I) were submitted to a curve fitting using the above power function [Eq. (1)]. Each fitting was characterized by its *Goodness-of-fit* (R^2) and the two constants (k and d). The distribution of the goodness-of-fit values across subjects is plotted in Fig. 1. Most values (N=71) are greater than 0.95 and no values smaller than 0.91. This indicates that the observed derived-band latencies as a function of frequency are described reasonably well by the proposed power function [Eq. (1)].

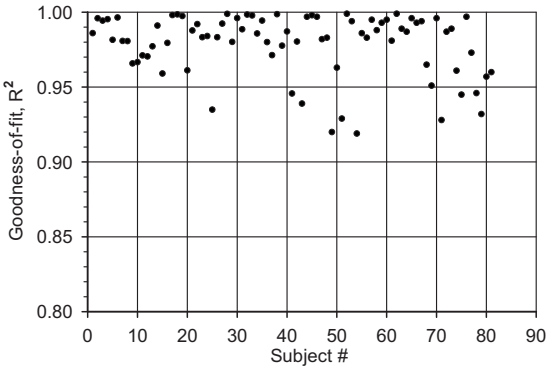


FIG. 1. Distribution across subjects of the goodness of fit (R^2) obtained by fitting the corrected derived-band latencies from each individual to a power-function model (N=81 Ss).

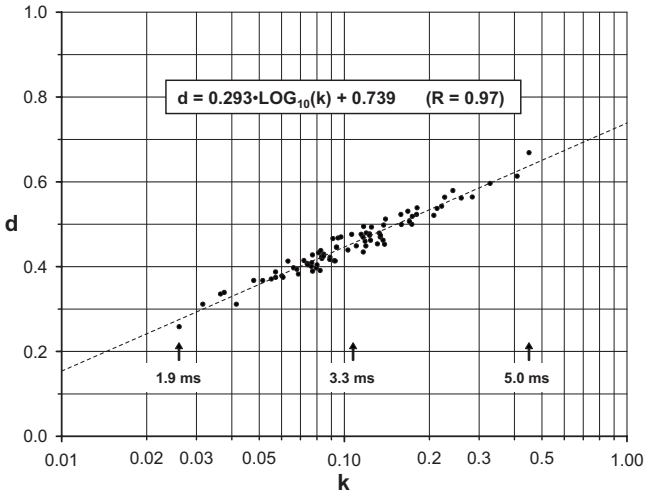


FIG. 2. Paired values of the two parameters, k and d, which define the power-function model [Eq. (1)] fitted to the corrected derived-band latencies from each individual (N=81 Ss). With k plotted on a logarithmic axis, there appears to be a linear relationship between d and k (indicated by the formula and the corresponding plotted broken line; R=0.97). The locations of the mean and the two end points of the distribution (k,d) are shown by the arrows together with the corresponding latency differences between the 5700 and 710 Hz derived bands.

The paired values of k and d are shown in Fig. 2, with k plotted on a logarithmic axis. The two fitting constants appear to have a logarithmic relationship as indicated by the regression line and the corresponding formula. Figure 2 demonstrates a significant interindividual spread in the latency-frequency functions. For each fitted power function, the latency delay between any two frequencies can be found from Eq. (1) by using the corresponding values of the two fitting constants k and d. Whereas the power function fitted to the mean values (Table I) has a latency delay of 3.3 ms between 5700 and 710 Hz, the individual power functions exhibit latency delays ranging from 1.9 to 5.0 ms.

C. Discussion

The combined dataset is based on data from both genders in a reasonably balanced way (43 females and 38 males). By analyzing the data from the females and males separately, the average derived-band latencies across the four frequency bands are 0.48 ms longer for the males than for the females. Further, the fitted power functions result in a latency delay between 5700 and 710 Hz which is 0.26 ms longer for the males than for the females. These gender differences have been examined more closely in the original publications by Don *et al.* (1993, 1998, 2005) and will not be discussed further.

The differences between the average values in datasets I and II are very small. This is quite surprising since the recordings for the two datasets are obtained seven years apart, with different groups of normal-hearing test subjects, with different earphones and by different examiners. However, the basic recording principles are identical. When the two sets of mean values are fitted to a power function which describes the latency-frequency relationships, the maximum difference between the two datasets appears at 710 Hz and is only 0.32 ms (within the frequency range 710–5700 Hz). At this frequency, the two fitted power functions deviate symmetrically (±0.16 ms) around the corresponding power function fitted to the mean values of the combined dataset. For all practical purposes, this means that there is only a marginal difference between using dataset I, dataset II, or the combined dataset.

Analysis of the derived-band latencies from all 81 normal-hearing test subjects reveals that the power-function model is able to describe each individual set of recorded latencies fairly accurately (goodness of fit, $R^2 > 0.91$). This indicates that the determined fitting constants (k and d) may be regarded as relevant descriptors of the fitted data. The paired k and d values are plotted in Fig. 2, which displays the interindividual variance of the latency-frequency information obtained from the derived-band ABRs.

In order to compensate for the cochlear traveling wave delay, chirp stimuli have previously been designed based on data representing the average cochlear delay in normal-hearing adults (e.g., Elberling *et al.*, 2007). The variance described above corresponds to a substantial variation (1.9–5.0 ms) in the latency delay between 5700 and 710 Hz across the group of 81 normal-hearing individuals. Even between the adjacent bands of 5700 and 2800 Hz, large differences in latency were reported earlier by Don *et al.* (1994). This implies that a chirp using a latency-frequency model based on the mean values of the combined dataset will not be optimal for all normal-hearing individuals. In an attempt to investigate this further, Don *et al.* (2009) formed three different Stacked ABRs from the derived-band responses underlying dataset II (N=39) and evaluated the amplitudes of the resulting waveforms. The different Stacked ABRs were formed by aligning the derived-band ABRs according to (1) the individual's peak latencies (standard procedure for the Stacked ABR), (2) the group mean latencies, and (3) the modeled latencies in accordance with a power-function model fitted to the average values of dataset II. The average peak-peak amplitudes for the three different Stacked ABRs were found to be 1014, 929, and 888 nV. This demonstrates that on the average the amplitude drops to about 0.88 (888/1014) or by 12% when modeled latencies are used instead of individual latencies. However, across the normal-hearing group in dataset II, the corresponding drop in amplitude ranges from −7% to 34%. Therefore, in normal-hearing subjects, we may expect response amplitudes up to about 35% lower than would have been produced by an individualized compensation of the cochlear delay.

III. DERIVED-BAND LATENCIES AT DIFFERENT STIMULUS LEVELS

A. Additional level information

In the study by Don *et al.* (1998), derived-band ABRs were obtained from normal-hearing subjects at a click level of 93 dB p.-p.e. SPL. The latencies from these recordings formed dataset I described above in Sec. II. In the same study, derived-band recordings were also obtained at other click levels and the results were reported from N=43 normal-hearing subjects at levels in 10 dB steps in the range from 93 to 53 dB p.-p.e. SPL (Don *et al.*, 1998). The wave V

TABLE II. Observed latencies (mean and standard deviation) for the 5700 and 1400 Hz derived-bands for different click levels. The observed latencies are corrected for the wave I-V delay (4.10 ms). The column to the right shows the calculated latency differences (delays) between the two derived bands

	Click level (dB p-p.e. SPL)	5700 Hz band latency (ms)		1400 Hz band latency (ms)		5700–1400 Hz difference (ms)	
		Mean	SD	Mean	SD	Mean	SD
	93	2.16	0.29	4.10	0.56	1.94	0.42
	83	2.43	0.29	4.32	0.62	1.89	0.48
$N=35$	73	2.75	0.34	4.75	0.75	2.00	0.55
	63	3.18	0.35	5.20	0.86	2.03	0.67
	53	3.68	0.54	6.06	0.93	2.38	0.73
$N=14$	43	4.25	0.58	6.56	1.18	2.31	0.91

latency in all four derived bands could not be measured at all click levels and in all subjects and therefore the latency delays (differences) between the derived bands at 5700 and 1400 Hz were the only ones that could be thoroughly analyzed. The results were described as follows: …"These data suggest that the delay between these two bands is, for most part, not dependent on stimulus level" (Don *et al.*, 1998).

The reported data, however, represented only a subset of the derived-band responses recorded from a total of $N=55$ normal-hearing subjects and for click levels all the way down to 43 dB p.-p.e. SPL. In the present study, we therefore reanalyze the obtained recordings in order to look more closely at the effect of level changes on the derived-band latencies. From the complete data pool, derived-band latencies were extracted from those normal-hearing subjects ($N=35$) from which the wave V in *both* the 5700 *and* the 1400 Hz band could be identified at *all* the levels 93, 83, 73, 63, and 53 dB p.-p.e. SPL. At the lowest level, i.e., 43 dB p.-p.e. SPL, wave V in both bands could only be identified in the recordings from $N=14$ subjects (a subgroup of the larger group of $N=35$ subjects). The Kolmogorov–Smirnov test of normality (Siegel, 1956) was performed on each dataset and indicated that none of the distributions deviates significantly from a Gaussian distribution described by the observed mean and standard deviations. The mean and standard deviations of the extracted latencies corrected for the wave I-V delay (4.1 ms) are shown in Table II and plotted in Fig. 3. The mean and standard deviations of the corresponding latency delays (differences) between the two bands are also shown in Table II.

Because the different datasets do not have the same variance (standard deviation, see Table II), comparison across stimulus level is done by means of nonparametric testing. By using Wilcoxon matched-pair signed-rank test (Siegel, 1956) on successive level pairs (including Bonferroni's correction for repeated testing—Hochberg and Tamhane, 1987) no difference was found in the latency delays between the two bands (5700 and 1400 Hz) at 93 versus 83 dB p.-p.e. SPL; 83 versus 73 dB p.-p.e.-SPL; and 73 versus 63 dB p.-p.e.SPL. Contrary to this, a significant difference ($p<0.01$) was found between the latency delays at 63 versus 53 dB p.-p.e. SPL. However, the mean difference between the delays at 63 and 53 dB p.-p.e. SPL is only 0.35 ms. Finally, no significant difference could be found in the latency delays at 53 versus 43 dB p.-p.e. SPL ($N=14$).

In Fig. 3, a curve (second-order polynomial) is fitted through the mean 5700 Hz derived-band latencies obtained in 10 dB steps within the range 43–93 dB p.-p.e. SPL. By extrapolation, it is estimated that this derived-band latency flattens off at about 110 dB p.-p.e. SPL with a value of 1.90 ms (see Fig. 3). This value is about 0.26 ms shorter than the corresponding value at 93 dB p.-p.e. SPL (2.16 ms).

B. Discussion

With the paired analysis, we cannot demonstrate statistically any change in the latency delay between the 5700 and 1400 Hz derived bands over the 30 dB range from

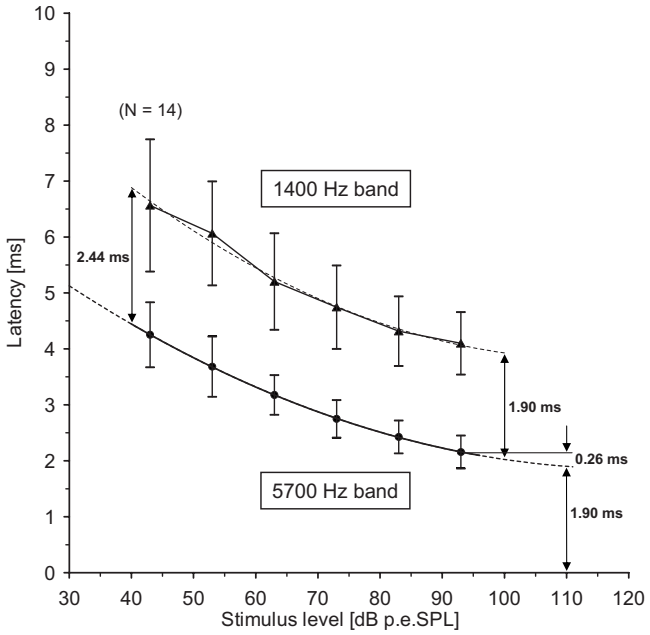


FIG. 3. Latencies (mean and ± one standard deviation) for the 5700 and 1400 Hz derived bands plotted as a function of click level (Table II). The mean values for the 5700 and the 1400 Hz derived bands are each fitted to a second-order polynomial, shown by the broken lines. The polynomial for the 5700 Hz band indicates that the latency in this band flattens off at about 110 dB p.-p.e.SPL with a value of 1.90 ms; this latency is 0.26 ms shorter than the observed mean value at 93 dB p.-p.e.SPL. The latency delay between the two bands estimated from the two fitted polynomials is 1.90 ms at 100 dB p.-p.e.SPL, and 2.44 ms at 40 dB p.-p.e.SPL.

63 to 93 dB p.-p.e. SPL. This is in agreement with the general description originally made by Don *et al.* (1998). However, at 53 dB p.-p.e. SPL, the latency delay increases significantly, which indicates that the derived-band responses from the two bands do not produce latency functions with a constant latency delay. It is conceivable that the latency function of the 1400 Hz band is more steeply sloping than that of the 5700 Hz band. A second-order polynomial is therefore also fitted to the mean latencies for the 1400 Hz band and compared to the fitted curve for the 5700 Hz band data, as shown in Fig. 3. The difference between the two fitted curves is 1.90 ms at 100 dB p.-p.e. SPL and 2.44 ms at 40 dB p.-p.e. SPL. If the two fitted curves represent the true underlying latency structure, the delay between the two bands only changes about 0.55 ms over a 60 dB range. Consequently, for all practical purposes, the two latency functions may be described as having a constant latency delay.

Derived band ABR's cannot readily be recorded at very high stimulus levels—mainly because these levels are perceived uncomfortably loud by the test subjects and because of risks of permanent damage from the high noise exposures. Derived band data are therefore not available for click levels higher than about 60 dB nHL (about 90–95 dB p.-p.e. SPL). From the latency data for the 5700 Hz band reported herein, we have therefore attempted to estimate the corresponding response latency at higher click levels. As shown in Fig. 3, the fitted curve indicates that the latency function flattens off at about 110 dB p.-p.e. SPL. The latency at this level could therefore be used as a reference in the comparison of derived-band and BM latency data observed at high stimulus levels (see Sec. IV).

IV. DERIVED-BAND LATENCIES AND THE TRAVELING WAVE

A. Observed derived-band latencies and the cochlear traveling wave delay

In order to compare the mean observed derived-band latencies from normal-hearing subjects [i.e., the combined dataset ($N=81$) in Table I], with the traveling wave delay, the recorded latencies have to be adjusted appropriately. All latencies in Table I include a subtracted value of 4.1 ms, corresponding to the average latency difference between ABR wave V and wave I (or ACAP wave N_1) found in a previous study (Elberling and Parbo, 1987).

No acoustical corrections were needed for the latencies in dataset I (Don *et al.*, 1998) in Table I. However, the latencies in dataset II (Don *et al.*, 2005) were corrected for the delay introduced by the tubing of the ER-2 earphone. Acoustical control measurements[3] demonstrated that the two earphones used in the two studies produce reasonably flat group delays across the frequency range 500–8000 Hz with average group delays of approximately 0.2 ms (after correcting the ER-2 for the sound tube delay). If correction is made for this average group delay the latency reference (0 ms) will approximate the time of arrival of the click at the tympanic membrane and the latencies would then represent traveling delays between the ear drum and the spiral ganglion cells in the cochlea—the site where the ACAP is generated (see, e.g., the Discussion in Kiang *et al.*, 1976).

The observed derived-band latencies (Table I) are not corrected for temporal delays caused by the synaptic gap. For this delay, we have chosen the value of 0.8 ms, which is similar to the value used by Eggermont (1979) and by Ruggero and Temchin (2007, p. 158).

Therefore, from the combined latency values in Table I, a value of 1.0 ms (=0.2+0.8) is finally subtracted. The corrected latencies are shown in Table I (last column—Corrected mean) and plotted in Fig. 7 together with a power function fitted to the four corrected data points.

How latencies of derived-band responses relate to delays of the cochlear traveling wave in humans—or the BM traveling wave—has recently been discussed by Ruggero and Temchin (2007). In that paper, the authors presented estimates of the signal-front delay and the weighted-average group delay of the BM as a function of location (characteristic frequency) in the cochlea of humans as well as of other species (see, e.g., Fig. 7 in Ruggero and Temchin, 2007). The estimates of the signal-front delay and the group delay suggested to be valid for humans at intense levels of stimulation are plotted in Figs. 6 and 7. From Fig. 7, it is quite clear that the corrected derived-band latencies deviate significantly from the suggested BM group delay.

B. Modeling derived-band responses

In order to look closer on this difference, a mathematical modeling of the derived-band responses is carried out by using a band-limited version of the formula originally proposed by Goldstein and Kiang (1958) to describe the formation of the ACAP N_1:

$$\mathrm{ACAP}_{ab}(t) = \sum_{n=a}^{b} \int_{-\infty}^{t} p_n(\tau) \cdot u_n(t-\tau) \cdot d\tau, \tag{2}$$

where $\mathrm{ACAP}_{ab}(t)$ is the derived-band ACAP from the frequency band between frequencies a and b; $p_n(t)$ is the firing probability of the nth nerve fiber represented by its poststimulus-time histogram, $\mathrm{PST}_n(t)$, and $u_n(t)$ is the unit response describing the contribution from the nth nerve fiber discharge to the activity at the remote recording electrode. In the original version of the formula, Goldstein and Kiang (1958) presumed the unit response to be the same for all nerve fibers. More recently, however, it has been suggested (e.g., Chertoff, 2004) that the shape of the unit response changes with the characteristic frequency of the nerve fibers. In this version of the formula, we only include contributions from nerve fibers within the frequency band in question. The right-hand side of the formula expresses the sum over the frequency band of the convolution of the firing probability function of each nerve fiber and the unit response. If the shape of the unit response can be regarded as being constant across the frequency band, the formula can be written as follows:

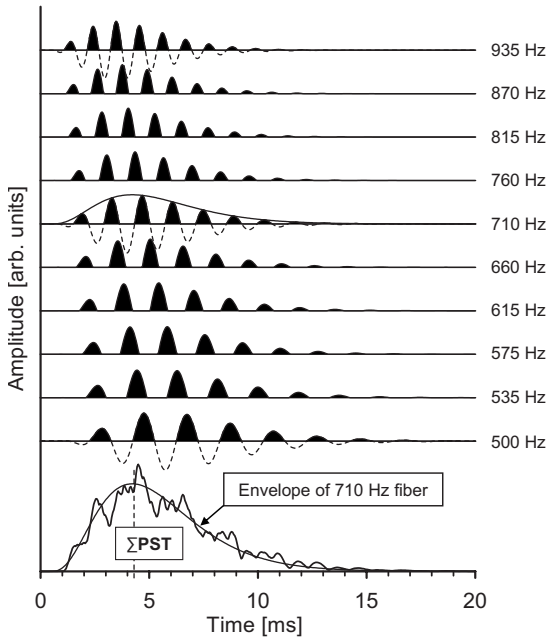


FIG. 4. (Top) Simulated activity from ten single nerve fibers, i.e., standard functions (de Boer, 1975) uniformly distributed on a logarithmic frequency scale of the low-frequency octave band. The neural activity or simulated PST histograms is indicated by the black areas. Each standard function is given a signal-front delay corresponding to the values estimated by Ruggero and Temchin (2007) and plotted in Figs. 6 and 7. (Bottom) The sum of the PST histograms (ΣPST) across the octave band (irregular curve). A scaled version of the temporal envelope of the standard function for the 710 Hz nerve fiber (the geometrical center frequency of the band) is plotted for comparison. The broken vertical line corresponds to the peak latency (4.3 ms) of the 710 Hz envelope.

$$\mathrm{ACAP}_{ab}(t) = \sum_{n=a}^{b} \mathrm{PST}_n(t) * u_n(t), \tag{3}$$

where $\Sigma_{n=a}^{b}\mathrm{PST}_n(t)$ is the sum of the PST histograms over the frequency band and $*$ is the convolution operator. The formula expresses that the derived-band response equals the sum of the PST histograms (in the following abbreviated ΣPST) filtered by the spectral characteristics of the unit response.

1. Modeling the ΣPST

To represent the ΣPST, we use a model of the auditory nerve fibers' temporal activity in response to click stimulation. Under the assumption of linearity, the *revcor* function is an approximation to the impulse response of the filtering process in each neural unit, e.g., Evans (1977) and de Boer and de Jongh (1978). Therefore, in a model of the ACAP in the cat, de Boer (1975) used a revcor template to model a *standard function*, for each fiber's impulse response (see Appendix A). A set of standard functions is shown in Fig. 5 (left).

If the click stimulus can be regarded as an approximation to a unit-impulse function within the frequency range from approximately 200 to 10 000 Hz, the standard function is the impulse response of the linear filter associated with a single nerve fiber. With this assumption, we have first attempted to model the derived-band activity from the low-frequency octave band (500–1000 Hz). Figure 4 (top) shows the simulated activity from ten single fibers uniformly distributed on a logarithmic frequency scale within the band. The neural activity—or simulated PST-histograms—is indicated by the black areas of the positive parts of the standard functions representing the activity of each nerve fiber. The standard functions are given signal-front delays ($t_{\mathrm{front}}=\alpha_f$; see Appendix A) in accordance with the values estimated by Ruggero and Temchin (2007), and plotted in Figs. 6 and 7. The sum of the PST histograms (ΣPST) across the octave band is shown in Fig. 4 (irregular curve, bottom). For comparison, the temporal envelope of the standard function for the nerve fiber with a characteristic frequency of 710 Hz, which corresponds to the geometrical center frequency of the band, is shown. This latter envelope is scaled in a way that enables visual comparison with the ΣPST curve. The broken vertical line indicates the peak latency of the 710 Hz response envelope (=4.3 ms). Thus the temporal waveform of the 710 Hz response envelope, including its peak latency, approximates the temporal waveform of the total firing probability of the group of single nerve fibers within the derived band. This corresponds to the suggestion by de Boer (1975, p. 1034) to replace the neural activity with its envelope when modeling the ACAP.

The envelope of the standard functions corresponding to the geometrical center frequencies of the four octave bands are subsequently used in Eq. (3) to represent the temporal activity to click stimulation—or the ΣPST—in each band. For the frequencies 710, 1400, 2800, and 5700 Hz, the four standard functions (and their envelopes) are shown in Fig. 5 (left). On each curve, the signal-front delay, the envelope-peak delay, and the weighted-average group delay are successively indicated and the values plotted in Fig. 6.

2. Modeling the unit response, $u_n(t)$

Different unit responses are reported in the literature—either derived theoretically, from masking experiments or from triggered averaging procedures in experimental animals as reviewed by Chertoff (2004). Since the reported unit responses deviate significantly from each other, we have used two models which are markedly different: the unit response originally proposed by de Boer (1975) (model 1) and the unit response proposed by Chertoff (2004) (model 2) (see, Appendix B). de Boer's unit response is the same for all nerve fibers, whereas Chertoff's unit response changes with characteristic frequency of the nerve fiber. The two unit responses are shown in Fig. 5 (middle) and are used for the convolution in Eq. (3).

3. Modeling of the derived-band ACAP

By convolving the envelope of the standard functions with the unit response (models 1 and 2), the modeled derived-band response waveforms are computed [Eq. (3)] and shown in Fig. 5 (right) with indication of the peak response latencies. It is apparent that the modeled derived-band response waveforms and their latencies depend significantly on the characteristics of the unit response. The resulting modeled latencies are plotted in Fig. 7. As can be seen, the corrected observed derived-band latencies deviate

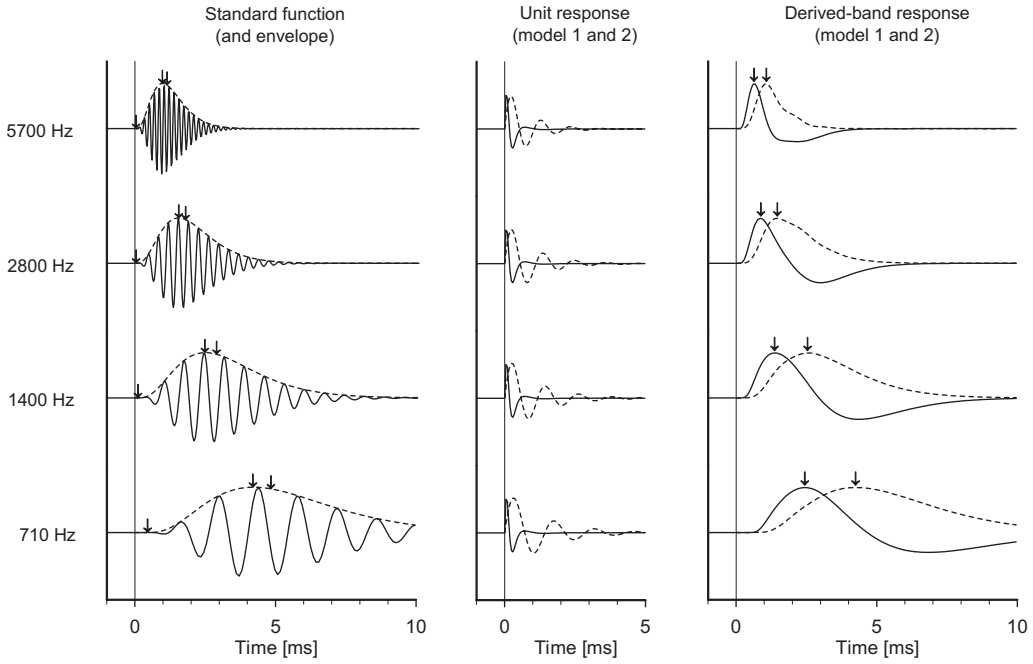


FIG. 5. (Left) Standard functions and envelopes (de Boer, 1975) corresponding to the center frequency of each of the four derived bands. The arrows indicate (left to right) the signal front delay, the envelope-peak delay, and the weighted-average group delay (see Appendix A); these latencies are plotted in Fig. 6. (Middle) Unit responses in accordance with models 1 (—) and 2 (- - -) (see Appendix B). (Right) Modeled derived-band responses obtained by convolution of the standard function envelopes (left) with each of the two unit responses (middle). The arrows indicate the peak latencies of the modeled derived-band responses; these latencies are plotted in Fig. 7.

significantly from the latencies of the modeled derived-band ACAPs using model 1, whereas they are more in agreement with those obtained using model 2.

C. Discussion

The high-pass masking paradigm, which has been used to record the derived-band ABR, presumes linearity (additivity of the high-pass masked responses as well as the corresponding derived-band responses). For the interpretation of the derived-band responses, it is also assumed that the fibers with characteristic frequencies inside the derived band respond fully and the fibers with characteristic frequencies out-

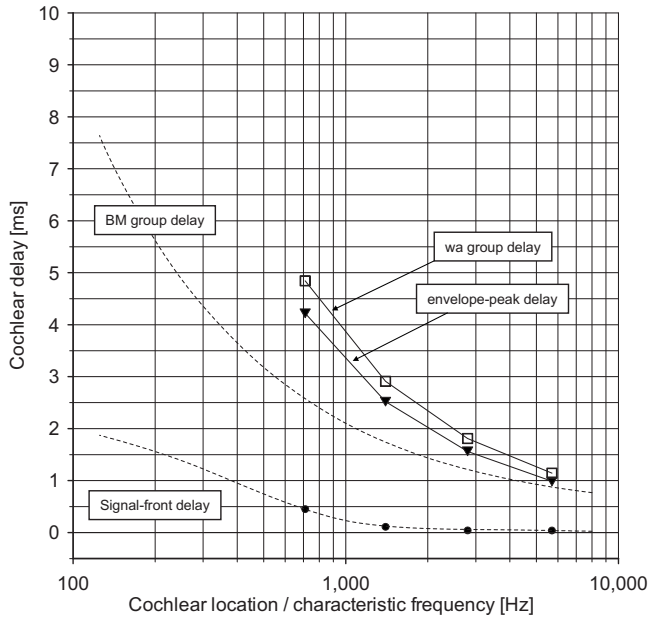


FIG. 6. Cochlear delay vs. cochlear location indicated by characteristic frequency. The broken lines show the signal front delay and the group delay of the BM estimated to be valid for humans at intense levels of stimulation by Ruggero and Temchin (2007, Fig. 7). The following delay values of the standard functions in Fig. 5 are shown: the signal front delay (●), the envelope-peak delay (▼), and the calculated weighted-average group delay (□).

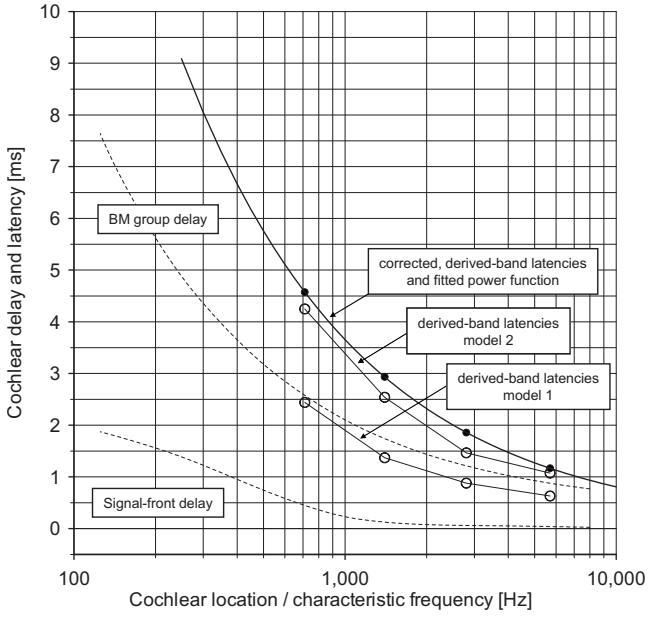


FIG. 7. Cochlear delay and latency vs cochlear location indicated by characteristic frequency. The broken lines show the signal front delay and the group delay of the BM estimated to be valid for humans at intense levels of stimulation by Ruggero and Temchin (2007, Fig. 7). The figure shows the corrected derived-band latencies (●) (Table I, last column) and the corresponding fitted power-function model (full line). The peak latencies (○) of the two sets of modeled derived-band responses (models 1 and 2; Fig. 5, right) are also included.

side the derived band do not respond at all. Prijs and Eggermont (1981) analyzed derived-band ACAPs from guinea pigs and concluded that the derived-band responses reflect the average behavior of activity from single fibers with characteristic frequencies corresponding to the center of the derived bands. In a high-pass masking experiment (Evans and Elberling, 1982), the activity in single auditory nerve fibers in the cat was studied. Apart from "remote masking" or lateral "two-tone" suppression between the click and the high-pass noise maskers which seemed to operate for low-frequency fibers, a general confirmation of the high-pass masking technique and therefore of the derived-band concept for the ACAP was supported. Therefore, we assume that the activity in each derived band reflects synchronized firing from nerve fibers with characteristic frequencies within the range between the two cutoff frequencies of the corresponding high-pass maskers.

The derived-band ABRs which produce datasets I and II (in Sect. II) were recorded at a moderate stimulus level i.e., 93 dB p.-p.e.SPL (~60 dB nHL). The observed latencies are therefore longer than would have been found at very intense click levels. However, we could attempt to calibrate the observed latencies for this level effect by subtracting 0.26 ms (Fig. 3) from the observed corrected latency values. However, this calibration would not bring the derived band latencies in line with the suggested BM group delay.

The data plotted in Fig. 3 indicate that the latency delay between the 5700 and 1400 Hz derived bands appears to be constant in the range between 93 and 63 dB p.-p.e.SPL (~60–30 dB nHL). For lower levels, i.e., 53 and 43 dB p.-p.e.SPL (~20 and 10 dB nHL), the latency delay becomes larger. For levels decreasing from 93 to 63 dB p.-p.e.SPL (from ~60 to 30 dB nHL), the derived-band latency-frequency curve in Fig. 7 would be shifted toward longer latencies. However, at 53 and 43 dB p.-p.e.SPL, the resulting latency-frequency curves would be shifted toward even longer latencies but more for the lower than for the higher bands, and the latency-frequency curves would therefore exhibit slightly more curvature than the 93 dB p.-p.e.SPL (~60 dB nHL) curve shown in Fig. 7.

In order to use the ABR wave V latency as a descriptor of the ACAP N_1 latency (ABR wave I), the same amount has been subtracted from the observed wave V latencies in all four derived bands (=4.1 ms). If the wave I-V difference in the low-frequency band were larger than in the high frequency band, a systematic error would be made. Such an error could then explain part of the observed differences between the experimental data and the modeled data and the suggested BM group delay. However, in a maturation study by Ponton *et al.* (1992), derived-band ABRs were recorded from newborns, infants, children, and adults in response to clicks presented at 91 dB p.-p.e.SPL (approximately 60 dB nHL). From the derived-band responses from the adults ($N=18$) the wave I-V delay was found to be constant across the derived bands (see Ponton *et al.*, 1992, Table I).

What is of more concern in the present context is the difference between the observed and the modeled derived-band latencies. It appears that this difference depends not only on the cochlear filter buildup time but also on the characteristics of the unit response used to model the derived-band response.

The revcor model is derived from the application of the reverse correlation method to the recordings from single auditory nerve fibers in the cat by, e.g., de Boer (1967), Evans (1977), and de Boer and de Jongh (1978). For low to moderate levels of the applied noise stimuli (up to 70 dB SPL per one-third octave) the revcor-functions are almost invariant (de Boer and de Jongh, 1978, p. 119) and are able to describe the filtering properties of the nerve fibers even at stimulus levels well above fiber saturation (Evans, 1977, p. 191). The click level used to record the derived-band ABRs in the present study is not excessive (93 dB p.-p.e. SPL ~60 dB nHL) and corresponds to a long-term level of about 67 dB SPL.[4] The click level complies therefore with the limitations given by de Boer and de Jongh (1978) and Evans (1977). However, regardless of the validity of this argument, we do not know with any certainty if the revcor function is an adequate first-order descriptor of the filtering process in the auditory nerve fibers for the data reported herein.

At very high click levels, we would expect the revcor functions to have significantly shorter rise times corresponding to the broader tuning of the neural units. The ΣPST curves in each octave band would then display earlier peak latencies than the ΣPSTs we have used for our modeling. These shorter latencies would be more in line with the BM group delay proposed by Ruggero and Temchin (2007).

The two unit responses used in the modeling represent different bandpass filters. The unit response proposed by de Boer (1975) corresponds to a filter with a slope of about 6 dB/octave in the frequency range of interest (below 700–900 Hz) and therefore acts almost as a differentiator. Consequently, the peak latencies of the modeled ACAPs (model 1) coincide with the maximum slope of the envelope of the standard functions (see Fig. 5). Contrary to this, the family of unit responses proposed by Chertoff (2004) corresponds to filters with a relatively flat response in the frequency range of interest. Consequently, the peak latencies of the modeled ACAPs (model 2) approach the actual peak latencies of the envelope of the standard functions (see Fig. 5).

Since neither the precise filtering process corresponding to the level of the click used to obtain the derived-band ABR nor the precise waveform of the unit responses of the involved nerve fibers are known, we are not able to provide a detailed explanation of the difference between the observed derived-band latencies and the suggested BM group delay. We therefore must agree with Ruggero and Temchin (2007, p. 158) that at present we have insufficient information to specify the relationship between the BM delays and derived-band latencies.

V. DESIGNING A CHIRP AND TESTING ITS EFFICIENCY

A. Designing a chirp

Individual nerve fiber responses or derived-band responses are obtained by measuring the activity from one individual nerve fiber or one group of nerve fibers at the time

and thus regarding the cochlea as a *tapped delay line*. Contrary to this, broadband ACAPs or ABRs are formed by the activity from the *whole* auditory nerve thus regarding the cochlea as a *two-port system*. For the design of chirp stimuli response latencies from specific locations along the BM obtained either from mechanical observations (von Bèkèsy, 1960; Ruggero and Temchin, 2007) or from responses from individual nerve fibers or group of nerve fibers have been used to characterize this two-port system (Fobel and Dau, 2004). In this way, estimates of the traveling wave delay have been used to characterize the group delay of the (linear) cochlear two-port system (e.g., Dau *et al.*, 2000; Wegner and Dau, 2002, and Elberling *et al.*, 2007). However, it is assumed that each derived band represents a neural activity pattern that arises from stimulation with frequency components *within* the frequency range of the derived band; this regardless of which location in the cochlea that actually generates the corresponding neural activity. Therefore, the derived-band latencies can be used to specify the group delay of a cochlear two-port system where the term frequency does not relate to position along the cochlear partition but to the input stimulus. This group delay then specifies how different groups of frequency components, for instance, a broadband click, and their resulting neural responses are delayed from the input to the output of the cochlear two-port system. The inverse group delay can subsequently be used to construct a chirp that attempts to compensate for the observed derived-band delay.

This approach has also been applied here. From the power function fitted to the mean latencies of the combined dataset (Table I, but offset to have zero delay at 10 000 Hz), a chirp was generated using the frequency-domain method described in detail by Elberling *et al.* (2007). Similarly, a standard 100 μs click was also generated by using a constant zero group delay.

The chirp used by Elberling *et al.* (2007) was generated from the mean latencies in dataset II. Since the differences are very small between dataset II and the combined dataset (Table I), the present chirp deviates only marginally from the one used previously (Elberling *et al.*, 2007)

The chirp and the click have different temporal characteristics. Therefore, if the efficiencies of the two stimuli are to be compared, it should be assured that they have identical amplitude spectra[5] and thus differ only in their phase spectra. The waveforms and amplitude spectra of the click and the chirp are shown in Fig. 8.

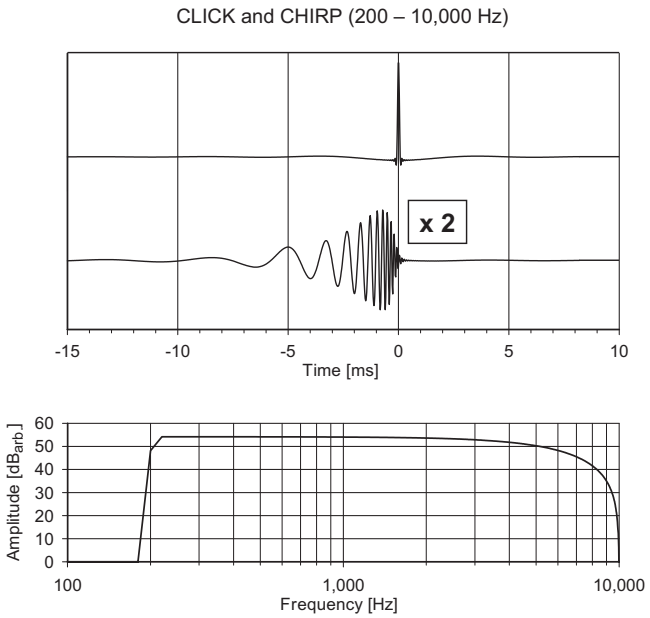


FIG. 8. (Top) The temporal waveform of the click and the chirp. The stimuli are band limited to the frequency range from 200 to 10 000 Hz. The amplitude of the chirp has been multiplied by 2. The two stimuli are temporally aligned so the phase delay at 10 000 Hz of both stimuli corresponds to 0 ms. (Bottom) The (identical) amplitude spectra of the two stimuli. The lower cutoff corresponds to −6 dB at 200 Hz and the higher cutoff corresponds to the falling spectrum of a 100 μs click, having its first null at 10 000 Hz (see endnote 5).

B. Testing the chirp against the click

1. Subjects

The clinical testing is part of a quality control procedure with CE-marked[6] equipment (used as intended), which complies with the general rules of the Central Ethical Committee and Danish Medicines Agency in Denmark. No written approvals are therefore needed. However, the following procedure was applied: Prior to testing, the purpose and procedure of the study were orally presented to each test subject by the experimenter. The experimenter then answered any questions the subject had regarding his/her participation. Finally, each subject read and signed an informed consent form.

The test group consisted of ten normal-hearing subjects (five females and five males) with ages 24–42 years. Normal hearing was defined as pure-tone thresholds of 10 dB HL or better for frequencies between 500 to 4000 Hz and 15 dB HL or better for 6000 and 8000 Hz. Both ears were tested sequentially on all test subjects.

2. Stimuli

The two test stimuli a 100 μs rarefaction click and the corresponding chirp (200–10 000 Hz) were applied to a set of ER-2 insert earphones. The stimuli were presented at a rate of 27/s (interstimulus interval=37 ms) at the levels 60 and 50 dB nHL.

3. ABR recordings

The present recordings were made with an Interacoustics Eclipse EP25 ABR-system®. The subjects were placed on a couch in an electrically shielded booth. ABRs were recorded differentially between electrodes applied high on the midfrontal area (F_z) and the ipsilateral mastoid (M_1 or M_2); an electrode on the lower midfrontal area (F_{pz}) was used as ground. The EEG was bandpass filtered from 100 to 3000 Hz using filter slopes of 12 dB/octave. ABRs were obtained using noise estimation techniques (Elberling and Don, 1984) and weighted averaging techniques (Elberling and Wahlgreen, 1985; Don and Elberling, 1994). These techniques reduce the destructive effects of episodic physiological background noise variation on the ABR average by weighting the average toward those blocks of sweeps with

TABLE III. Observed wave-V peak-peak amplitudes and peak latencies (mean and standard deviation) of the recorded click and chirp ABRs at the stimulus levels 60 and 50 dB nHL ($N=20$ ears). The column to the right shows the calcualted ratios between the chirp and the click ABR-amplitudes.

Stimulus level (dB nHL)	Click				Chirp					
	Amp. p-p (nV)		Latency (ms)		Amp. p-p (nV)		Latency (ms)		Amp. ratio	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
60	331	69	6.12	0.22	501	104	4.62	0.66	1.54	0.32
50	305	58	6.55	0.41	531	132	5.43	0.66	1.78	0.47

low estimated background noise. For the purpose of collecting the data presented here, averaging was continued in each run until the estimated residual background noise level was 40 nV (rms) or lower. For all test subjects and all recordings, an average residual background noise level of 34.6 nV was obtained after 3150 sweeps.

4. Wave V parameters

The Wave V peak-to-trough amplitude and peak latency were identified and measured automatically from the recordings. To compensate for the acoustical delay in the sound tubing of the ER-2 earphones, 0.86 ms was subtracted from all measured latency values. In six of the recordings from one subject, significant postauricular muscle (PAM) response precluded the automatic amplitude measurements because the trough following the wave V was obscured by the PAM activity. In these cases, the location of the trough was found by judgments made by two independent observers. The Kolmogorov–Smirnov test of normality (Siegel, 1956) was performed on the values from the data from each condition. The results indicated that none of the data distributions deviates significantly from a Gaussian distribution described by the observed mean and standard deviations. Thus, it was assumed that all the data reported here are approximately normally distributed, justifying the use of a normal parametric description of the data. However, because the samples from the different conditions do not have the same variance, we have applied nonparametric testing for the statistical analyses: The Wilcoxon matched-pair signed-rank test for analysis of paired data and the Mann–Whitney U test for the analysis of unpaired data (Siegel, 1956).

5. Results

The mean and standard deviations of peak-peak amplitude and peak latency for all four sets of recordings (from both ears in ten subjects) are shown in Table III. At both 60 and 50 dB nHL, there are marked differences between the click and chirp data: the chirp amplitudes are significantly larger than the corresponding click amplitudes ($p<0.001$, Wilcoxon) and the chirp latencies significantly shorter than the click latencies ($p<0.001$, Wilcoxon). Whereas the *relative* standard deviations for the chirp amplitudes are about the same as those for the corresponding click amplitudes, the standard deviations for the chirp latencies are significantly larger than those for the corresponding click latencies [$p<0.001$ (60 dB) and $p<0.05$ (50 dB), F-test].

At both 60 and 50 dB nHL, the ratio between the chirp amplitudes and the click amplitudes are calculated and the mean and standard deviations shown in Table III. At 60 dB nHL, the mean ratio is 1.54 and at 50 dB nHL it is 1.78. The two observed cumulative distributions of the amplitude ratios are plotted together with their estimated Gaussian distributions in Fig. 9.

In order to study the ABR waveform morphology, grand averages are constructed as follows: First, each recording is time shifted so the latency of wave V coincides with the mean latency for the actual recording condition (mean laten-

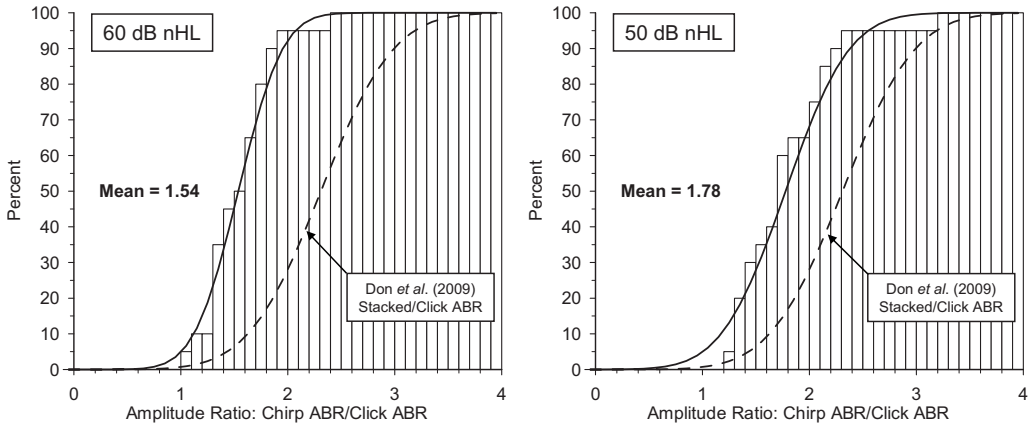


FIG. 9. Cumulative distribution of the ratios between the chirp and the click ABR amplitudes ($N=20$ ears) obtained at 60 dB nHL (left, mean value=1.54) and at 50 dB nHL (right, mean value=1.78). Both the observed distribution of the individual amplitude ratios and the corresponding continuous Gaussian distribution using the observed mean and standard deviation (Table III) are shown. The continuous Gaussian distribution of the ratios between the modeled Stacked and the click ABR amplitudes obtained at 60 dB nHL by Don *et al.* 2009 [Fig. 1(c)] are shown for comparison (broken line).

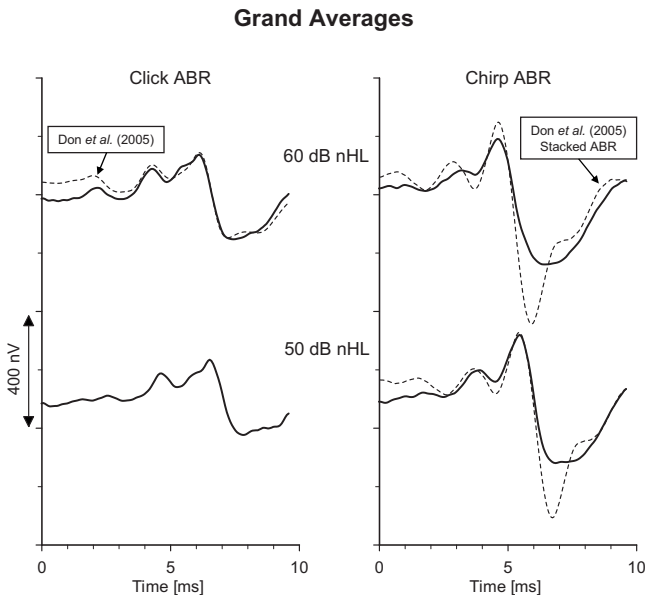


FIG. 10. Grand average waveforms of click ABR (left) and chirp ABR (right) obtained at 60 and 50 dB nHL ($N=20$ ears). The corresponding grand averages of the click ABR and the modeled Stacked ABR obtained from the original recordings at 60 dB nHL by Don *et al.* (2005) are shown for comparison (broken lines): left, click ABR; right, modeled Stacked ABR. These latter waveforms are scaled and time shifted as described in detail in the text.

cies for each stimulus type and level, Table III). Next, the average across each set of 20 time-adjusted recordings is calculated. The four grand average waveforms are shown in Fig. 10. For comparison, the grand averages of the click ABR and Stacked ABR[7] obtained at 60 dB nHL by Don *et al.* (2005) [dataset II ($N=39$), Sec. II] are constructed in a similar way, as described above. In Fig. 10, these latter grand averages are plotted so their peak latencies coincide with those from the present study. Further, the two grand averages (click and Stacked ABR, Don *et al.*, 2005) are multiplied by a scaling factor=0.88 which enables the grand average click ABR to be presented with the same peak-peak amplitude as the grand average click ABR from the present study.[8] Since the waveform of the time and amplitude-adjusted grand average click ABR (Don *et al.*, 2005) resembles that from the present study ($R^2=0.90$), the click ABRs at 60 dB nHL serve as a common reference and the chirp ABR can then meaningfully be compared with the Stacked ABR. At 60 dB nHL the waveform of the Stacked ABR is slightly different from the chirp ABR ($R^2=0.80$) and also significantly larger, but at 50 dB nHL the two waveforms become more similar ($R^2=0.85$).

C. Discussion

As found in other studies (e.g., Dau *et al.*, 2000, Fobel and Dau, 2004, and Elberling *et al.*, 2007), we find the chirp ABR amplitude to be larger than the corresponding click ABR amplitude. The chirp/click amplitude ratio is significantly lower at 60 than at 50 dB nHL (1.54 versus 1.78, $p<0.01$, Wilcoxon), because when the level of stimulation drops from 60 to 50 dB nHL the click ABR amplitude decreases whereas the chirp ABR increases. This finding is in agreement with the findings by Fobel and Dau (2004, Fig. 4), and from their results the amplitude ratios 1.42 (60 dB SL) and 2.05[9] (50 dB SL) can be estimated (for their M-chirp).

It should be noted that the observed (shorter) latencies of the chirp ABRs have significantly larger standard deviations than the corresponding latencies of the click ABRs. However, the analysis of the chirp data demonstrates a very high correlation between the latencies at 60 and 50 dB nHL ($R=0.95$), whereas for the click the corresponding correlation is much lower ($R=0.80$). The higher the correlation, the more the observed variance is due to individual differences than due to random fluctuations. The result therefore seems to indicate that the chirp ABR systematically emphasizes individual characteristics—at least for the two stimulus levels used herein.

In the study by Elberling *et al.* (2007), the terms "input compensation" and "output compensation" were introduced: Input compensation refers to the use of a chirp stimulus which attempts to compensate for the traveling wave delay at the *input* to the cochlea and output compensation refers to the application of the Stacked ABR which attempts to compensate for the traveling wave delay at the *output* from the cochlea. In the study by Don *et al.* (2009), the click ABR at 60 dB nHL and the corresponding Stacked ABR were evaluated in the recordings from 39 normal-hearing subjects and three different Stacked ABRs were formed (also described above in Sec. II C). One of these—called the modeled Stacked ABR—was made in an attempt to simulate the effect of using a chirp stimulus which in principle would compensate for the traveling wave delay similar to the Stacked ABR when based on the same average latency-frequency model. The mean ratio of the modeled Stacked ABR amplitudes to the corresponding click ABR amplitudes across the group of 39 normal-hearing subjects was found to be 2.27 [sd=0.50 (Don *et al.*, 2009, Fig. 1(c))]. For comparison with the present findings, the corresponding Gaussian cumulative distribution is shown in the two graphs in Fig. 9.

At 60 dB nHL, the chirp/click amplitude ratio is significantly lower than the corresponding modeled stacked/click amplitude ratio ($p<0.001$, Mann–Whitney). This means that the gain in amplitude relative to the click ABR is higher for the modeled Stacked ABR than for the chirp ABR. For both compensation methods, it can be observed that no recordings produce a ratio <1.0. However, the distribution of the chirp/click ratio is much steeper than the modeled stacked/click ratio, corresponding to a significantly lower observed standard deviation (Table III) of 0.32 versus 0.50 ($p<0.02$, F-test), but the relative standard deviations are about the same (0.21 versus 0.22).

At 50 dB nHL, the modeled stacked/click amplitude ratio (obtained at 60 dB nHL) is also higher than the corresponding chirp/click amplitude ratio. However, here the two distributions appear almost equally steep corresponding to an observed standard deviation (Table III) of 0.47 versus 0.50 (NS, F-test), and the relative standard deviations are only slightly different (0.26 versus 0.22).

In Fig. 10, the corresponding waveforms are evaluated by means of the time-adjusted grand averages. All the recordings by Don *et al.* (2005) are scaled down with the pur-

pose of presenting the two grand average click ABRs at 60 dB nHL with the same peak-peak amplitude. By using the click ABR as an anchor point or common reference for the two sets of recordings, the waveforms of the modeled Stacked ABR and the chirp ABRs can be compared. The differences in the amplitude ratios as found above repeat itself in the waveforms. At 60 dB nHL, the waveform of the chirp ABR has some similarities with the modeled Stacked ABR but is much smaller. However, at 50 dB nHL, the two waveforms become more similar and more equal in magnitude.

Therefore, at 50 dB nHL, the efficiency of input compensation (chirp ABR) and of output compensation (the *estimated* modeled Stacked ABR) in generating larger ABRs appear to be comparable when obtained using similar recording conditions.

That the chirp is more efficient at 50 dB nHL than at 60 dB nHL could be related to the spread of excitation by the individual frequency components of the chirp. At lower levels of stimulation, each frequency component excites only a restricted area of the cochlea, but at higher levels the excitation spreads—especially toward the base of the cochlea (upward spread of excitation). Each frequency component of the chirp arrives at the cochlea with one specific delay, which is designed to compensate for the cochlear delay at the location with a characteristic frequency corresponding to the frequency component in question. If each frequency component at higher levels excites a broader area of the cochlea, the effect will be a desynchronization, because each location will now be excited by a broader range of frequency components each arriving at a different point in time. If this explanation is correct, there might be an upper level of stimulation beyond which the chirp no longer will be more effective than the click—at least not in normal-hearing individuals.

VI. SUMMARY AND CONCLUSION

There are only small differences between the two datasets (I and II) that are used to create the combined dataset describing the latency-frequency relationship of derived-band ABRs obtained at 60 dB nHL in normal-hearing individuals.

The proposed power-function model is an adequate descriptor of the combined dataset as well as the latency-frequency relationship in the individual.

There is a large interindividual variance in the delay between the derived-band ABRs across normal-hearing individuals, corresponding to a range from 1.9 to 5.0 ms between the 5700 and 710 Hz bands.

The observed variance indicates that a chirp constructed from the mean values in the combined dataset may cause reductions of up to 35% in response amplitude relative to an individualized compensation of the cochlear delay in normal-hearing subjects.

Inferred from derived-band ABRs obtained at different click levels there seems to be no major effect on the relative cochlear delay ($<$0.55 ms—between 5700 and 1400 Hz) over a 60 dB range of stimulation ($\sim$10–70 dB nHL).

Extrapolation of the average latency values obtained for click levels from 10 to 60 dB nHL indicates that the 5700 Hz derived-band latency function flattens off at $\sim$80 dB nHL (110 dB p.-p.e.SPL).

The observed latency of derived-band ABRs depends not only on the cochlear filter buildup time but also on the waveform (or frequency spectrum) of the unit response, which characterizes the contribution to the activity at the remote recording electrode of the discharge in single nerve fibers.

In humans neither the shape of the cochlear filter response nor the waveforms of the unit responses are known in detail. Therefore, the relationship between the derived-band latencies and the suggested group delay of the BM cannot be specified.

Chirp ABRs generate higher response amplitudes than the corresponding click ABRs. However, the gain in amplitude is lower at 60 dB nHL than at 50 dB nHL where it approaches a factor of 2.

The drop in response amplitude of the chirp ABR over the click ABR for higher levels of stimulation in normal-hearing individuals implies that there might be an upper level of stimulation beyond which the chirp is no longer more effective than the click.

From the latency behavior of the chirp ABR, it is suggested that the chirp emphasizes individual characteristics more than the corresponding click.

Chirp ABRs are lower in magnitude than Stacked ABRs at 60 dB nHL. However, at 50 dB nHL, the magnitude of chirp ABRs seems to approach that of modeled Stacked ABRs, and the two methods of compensation appear therefore to be equally efficient.

ACKNOWLEDGMENTS

The authors want to thank Johannes Callø, M.Sc., Interacoustics, Denmark, for his management of the test subjects and the collection of the electrophysiological data. The authors also want to acknowledge Søren Laugesen, Ph.D., "Eriksholm," Oticon A/S, Denmark, for his constructive criticism and support to parts of the work presented herein. The work is supported in part by Grant No. 1R01 DC03592 (P.I. Manuel Don) from NIDCD at NIH.

APPENDIX A: REVCOR FUNCTIONS

In a model of the ACAP in the cat, de Boer (1975) used a revcor template to model a *standard function*, $h(t)$, for each fiber's impulse response:

$$h(t) = \mathrm{env}(t) \cdot \cos(2\pi \cdot f_C \cdot t + \varphi_{f_C}), \tag{A1}$$

where t is the time in seconds, f_c is the fiber's characteristic frequency in Hertz, and env(t) is the temporal envelope of $h(t)$ defined as follows:

$$\mathrm{env}(t) = 0 \quad \text{for } t < \alpha_{f_C},$$

$$\mathrm{env}(t) = [(t-\alpha_{f_C})/\beta_{f_C}]^{(\gamma-1)} \cdot e^{-[(t-\alpha_{f_C})/\beta_{f_C}]} \quad \text{for } t \geq \alpha_{f_C}, \tag{A2}$$

where α_{f_C} is a delay time dependent on f_c, β_{f_C} is a decay constant$=0.0975 f_C^{-0.663}$, γ is a degree of complexity of the associated filter (chosen by de Boer to be=4.0), and φ_{f_C} is the phase of the "carrier" or characteristic frequency (chosen by de Boer (1975) to be be$=-\alpha_{f_C} \cdot 2\pi \cdot f_C$).

Several different delays are associated with this standard function: the *signal-front delay* t_{front} $(=\alpha_f)$, the envelope-peak delay t_{peak}, and the *weighted-average group delay* t_{group}, (see Goldstein *et al.*, 1971, Ruggero 1980, 1994). In a linear system, the weighted-average group delay corresponds to the center of gravity of the impulse response energy (see, e.g., Goldstein *et al.*, 1971). A corresponding numerical calculation demonstrates that for a signal-front delay=0, de Boer's standard function, Eq. (A1), gives a weighted-average group delay which is approximately 17% longer than the envelope-peak delay, i.e., $t_{\mathrm{group}}=1.17 t_{\mathrm{peak}}(\alpha_f=0)$.

APPENDIX B: UNIT RESPONSES

Model 1: Unit response proposed by de Boer

This unit response is constant across nerve fibers and is defined in the *s*-domain through its Laplace transform:

$$U_n(s) = \frac{s^2}{(1+s)\cdot(1+\xi\cdot s+s^2)}, \tag{B1}$$

where the constant $\xi=1.2$ and $s=1$ corresponds to a frequency of 1000 Hz. The corresponding time function is rather complex and does not have a zero-mean value. Therefore, we have used the following mathematical approximation:

$$u_n(t) = e^{-at}\sin(2\pi\cdot(b\cdot e^{-ct})\cdot t), \tag{B2}$$

where the constants are given the values (in s^{-1}): $a=2900$, $b=3600$, and $c=1250$, which ensures the unit response a zero-mean value.

Model 2. Unit response proposed by Chertoff

This unit response changes with the characteristic frequency of the nerve fiber, and consists of a damped sinusoidal function given as follows:

$$u_n(t) = e^{-kt}\sin(2\pi f\cdot t), \tag{B3}$$

where the damping factor k is in s^{-1} and the oscillation frequency f in Herz. From the information provided by Chertoff (2004, Figs. 4 and 5), we have approximated the two constants with the following values as a function of the characteristic frequency: (*characteristic frequency*, f, k)$\Rightarrow$(*710*, 700, 750), (*1400*, 850, 900), (*2800*, 900, 1050), and (*5700*, 950, 1250).

[1]Don *et al.* (1998) reported data from 43 normal-hearing subjects. However, since the dataset from one male was not complete, it was excluded in the present analysis.

[2]The THD 50P earphone was calibrated on the artificial ear: IEC 60318-3 (1998), (i.e., Brüel & Kjær 4152 with a 6-cc coupler) and the click level was 93 dB p.-p.e. SPL, which corresponded to about 63 dB nHL. The ER-2 insert earphone was calibrated in the artificial ear: IEC 60318-5 (2006), (i.e., Brüel & Kjær 4152 with a 2-cc coupler) and the click level was 82 dB p.-p.e. SPL, which corresponded to about 60 dB nHL.

[3]For these measurements the TDH 50P earphone was coupled to the artificial ear: IEC-60318-3 (1998), (Brüel & Kjær 4152 with a 6-cc coupler), and the ER-2 insert earphone was coupled to the occluded ear simulator: IEC 60711 (1981), (i.e., Brüel & Kjær 4157). The impulse responses of the two earphones including the couplers were measured by a PULSE measurement system (Brüel & Kjær 3560 C). Subsequently, the amplitude response, phase response, phase delay, and group delay were calculated for each earphone.

[4]Numerical calculation reveals that a 100 μs click at a rate of 45/s gives a rms to peak-peak amplitude ratio of 0.052 ($\sim$−26 dB) in the frequency range 200–10 000 Hz. Therefore, 93 dB p.-p.-p.e. SPL corresponds roughly to 67 dB SPL.

[5]Electrically both stimuli are limited to the frequency range 200–10 000 Hz. Within this range, the amplitude-frequency characteristic corresponds to the $|\sin(x)|/|x|$-function of the 100 μs click having its first spectral null at 10 000 Hz.

[6]The CE marking is a mandatory conformity mark for most products in the European Economic Area.

[7]The individual stacked ABRs are formed by using the modeled latencies instead of the individual latencies of the derived band ABRs, as described by Don *et al.* (2009). However, in the study of Don *et al.* (2005), recordings are only available at 60 dB nHL. An *estimate* of the Stacked ABR at 50 dB nHL is obtained from the stacked ABR at 60 dB nHL by reducing its magnitude by 10% corresponding to the observed reduction in amplitude of the click ABR from 60 and 50 dB nHL (Table III).

[8]There are two important differences in the recording conditions used in the two studies: (1) electrode montage, vertex (C_z)—mastoid (M_1 and M_2) versus high forehead (F_z)—mastoid (M_1 and M_2), and (2) stimulus rate, 45/s versus 27.1/s. The scaling factor compensates (at least in part) for the corresponding differences in response amplitude.

[9]In the referred study, the click and the chirp did not have the same amplitude spectrum because the chirp was about 3 dB larger than the click above approximately 4 kHz (Fobel and Dau, 2004, Fig. 2). Artificially, this may have increased the relative amplitudes of the chirp ABR and thereby the corresponding amplitude ratios.

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (**1971**). "Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects," J. Acoust. Soc. Am. **49**, 1131–1139.

Chertoff, M. E. (**2004**). "Analytic treatment of the compound action potential: Estimating the summed post-stimulus time histogram and unit response," J. Acoust. Soc. Am. **116**, 3022–3030.

Dau, T., Wagner, O., Mellert, V., and Kollmeier, B. (**2000**). "Auditory brainstem responses with optimized chirp signals compensating basilar membrane dispersion," J. Acoust. Soc. Am. **107**, 1530–1540.

de Boer, E. (**1967**). "Correlation studies applied to the frequency resolution of the cochlea," J. Aud Res. **7**, 209–217.

de Boer, E. (**1975**). "Synthetic whole-nerve action potentials for the cat," J. Acoust. Soc. Am. **58**, 1030–1045.

de Boer, E. (**1980**). "Auditory physics. Physical principles in hearing theory I," Phys. Rep. **62**, 87–174.

de Boer, E., and de Jongh, H. R. (**1978**). "On cochlear encoding: potentials and limitations of the reverse-correlation technique," J. Acoust. Soc. Am. **63**, 115–135.

Don, M., and Eggermont, J. J. (**1978**). "Analysis of click-evoked brainstem potentials in man using high-pass masking," J. Acoust. Soc. Am. **63**, 1084–1092.

Don, M., Eggermont, J. J., and Brackmann, D. E. (**1979**). "Reconstruction of the audiogram using brainstem responses and high-pass noise masking," Ann. Otol. Rhinol. Laryngol. Suppl. **57**, 1–20.

Don, M., and Elberling, C. (**1994**). "Evaluating residual background noise in human auditory brainstem responses," J. Acoust. Soc. Am. **96**, 2746–2757.

Don, M., Elberling, C., and Maloff, E. (**2009**). "Input and output compensation for the cochlear travelling wave delay in wide-band ABR recordings: Implications for small tumor detectiona ," J. Am. Acad. Audiol **20**(2).

Don, M., Kwong, B., and Tanaka, C. (**2005**). "A diagnostic test for Meniere's disease and cochlear hydrops: Impaired high-pass noise masking of auditory brainstem response," Otol. Neurotol. **26**, 711–722.

Don, M., Masuda, A., Nelson, R., and Brackmann, D. (**1997**). "Successful

detection of small acoustic tumors using the stacked derived-band auditory brain stem response amplitude," Am. J. Otol. **18**, 608–621.
Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1993**). "Gender differences in cochlear response time: An explanation of gender amplitude differences in the unmasked auditory brain-stem response," J. Acoust. Soc. Am. **94**, 2135–2148.
Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1994**). "Auditory brainstem response (ABR) peak amplitude variabllity reflects individual differences in cochlear response times," J. Acoust. Soc. Am. **96**, 3476–3491.
Don, M., Ponton, C. W., Eggermont, J. J., and Kwong, B. (**1998**). "The effects of sensory hearing loss on cochlear filter times estimated from auditory brainstem response latencies," J. Acoust. Soc. Am. **104**, 2280–2289.
Eggermont, J. J. (**1979**). "Compound action potentials: tuning curves and delay times," Scand. Audiol. Suppl. **9**, 129–139.
Elberling, C., and Don, M., (**1984**). "Quality estimation of averaged auditory brainstem responses," Scand. Audiol. **13**, 187–197.
Elberling, C., Don, M., Cebulla, M., and Stürzebecher, E. (**2007**). "Auditory steady-state responses to chirp stimuli based on cochlear traveling wave delay," J. Acoust. Soc. Am. **122**, 2772–2785.
Elberling, C., and Parbo, J. (**1987**). "Reference data for ABR's in retrocochlear diagnosis," Scand. Audiol. **16**, 49–55.
Elberling, C., and Wahlgreen, O. (**1985**). "Estimation of auditory brainstem responses, ABR, by means of Bayesian inference," Scand. Audiol. **14**, 89–96.
Evans, E. F. (**1977**). "Frequency selectivity at high signal levels of single units in cochlear nerve and nucleus," in *Psychophysics and Physiology of hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 185–192.
Evans, E. F., and Elberling, C. (**1982**). "Location-specific components of the gross cochlear action potential," Audiology **21**, 204–227.
Fobel, O., and Dau, T. (**2004**). "Searching for the optimal stimulus eliciting auditory brainstem responses in humans," J. Acoust. Soc. Am. **116**, 2213–2222.
Goldstein, J. L., Baer, T., and Kiang, Nelson Y. S. (**1971**). "A theoretical treatment of latency, group delay, and tuning characteristics for auditory-nerve responses to clicks and tones," in *The physiology of the auditory system*, edited by M. B. Sachs (National Educational Consultants, Baltimore MD), pp. 133–141.
Goldstein, Jr., M. H., and Kiang, Nelson Y. S. (**1958**). "Synchrony of neural activity in electric responses evoked by transient acoustic stimuli," J. Acoust. Soc. Am. **30**, 107–114.
Hochberg, Y., and Tamhane, A. C. (**1987**). *Multiple Comparison Procedures* (Wiley, New York).
IEC 60318-3. (**1998**). "Electroacoustics—Simulators of human head and ear—Part 3: Acoustic coupler for the calibration of supra-aural earphones used in audiometry" (International Electrotechnical Commission, Geneva, Switzerland).
IEC 60318-5. (**2006**). "Electroacoustics—Simulators of human head and ear—Part 5: 2 cm^3 coupler for the measurement of hearing aids and earphones coupled to the ear by means of ear inserts" (International Electrotechnical Commission, Geneva, Switzerland).
IEC 60711. (**1981**). "Occluded ear simulator for the measurement of earphones coupled to the ear by ear inserts" (International Electrotechnical Commission, Geneva, Switzerland).
Kiang, N-Y. S., Moxon, E. C., and Kahn, A. R. (**1976**). "The relationship of gross potentials recorded from the cochlea to single unit activity in the auditory nerve," in *Electrocochleography*, edited by R. J. Ruben, C. Elberling, and G. Salomon (University Park, Baltimore), pp. 95–115.
Lütkenhöner, B., Kauffmann, G., Pantev, C., and Ross, B. (**1990**). "Verbesserung der Synchronisation auditorisch evozierter Hirnstammpotentiale durch Verwendung eines die cochleären Laufzeitunterschiede kompensierenden Stimulus (Increased synchronization of the auditory brainstem response obtained by a stimulus which compensates for the cochlear delay)," Arch. Otolaryngol. **2**, 157–159.
Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). "Latency of auditory brain-stem repsonses and otoacoustic emissions using tone-burst stimuli," J. Acoust. Soc. Am. **83**, 652–656.
Parker, D. J., and Thornton, A. R. D. (**1978**). "Frequency specific components of the cochlear nerve and brainstem evoked responses of the human auditory system," Scand. Audiol. **7**, 53–60.
Ponton, C., Eggermont, J. J., Coupland, S. G., and Winkelaar, R. (**1992**). "Frequency-specific maturation of the eight nerve and brain-stem auditory pathway: Evidence from the derived auditory brain-stem responses (ABRs)," J. Acoust. Soc. Am. **91**, 1576–1586.
Prijs, V. F., and Eggermont, J. J. (**1981**). "Narrow-band analysis of compound action potentials for several stimulus condition in the guinea pig," Hear. Res. **4**, 23–41.
Ruggero, M. A. (**1980**). "Systematic errors in indirect estimates of basilar membrane travel times," J. Acoust. Soc. Am. **67**, 707–710.
Ruggero, M. A. (**1994**). "Cochlear delays and traveling waves: Comments on 'Experimental look at cochlear mechanics'," Audiology **33**, 131–142.
Ruggero, M. A., and Temchin, A. N. (**2007**). "Similarity of traveling wave delays in the hearing organs of humans and other tetrapods," J. Assoc. Res. Otolaryngol. **8**, 153–166.
Shore, S. E., and Nuttall, A. L. (**1985**). "High-synchrony cochlear compound action potentials evoked by rising frequency-swept tone bursts," J. Acoust. Soc. Am. **78**, 1286–1295.
Siegel, S. (**1956**). *Nonparametric Statistics for the Behavioural Sciences*, McGraw-Hill, London.
Stürzebecher, E., Cebulla, M., Elberling, C., and Berger, T. (**2006**). "New efficient stimuli for evoking frequency-specific auditory steady-state responses," J. Am. Acad. Audiol **17**, 448–461.
Teas, D. C., Eldredge, D. H., and Davis, H. (**1962**). "Cochlear response to acoustic transients: An interpretation of whole-nerve action potentials," J. Acoust. Soc. Am. **34**, 1438–1459.
von Békésy, G. (**1960**). *Experiments in Hearing*, McGraw-Hill, New York.
Wegner, O., and Dau, T. (**2002**). "Frequency specificity of chirp-evoked auditory brain stem responses," J. Acoust. Soc. Am. **111**, 1318–1329.

Representation of the spectral dominance region of pitch in the steady-state temporal discharge patterns of cochlear nucleus units

William P. Shofner[a)]
Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, Indiana 47405

(Received 14 July 2008; revised 11 August 2008; accepted 12 August 2008)

Single-unit responses to infinitely iterated rippled noise and wideband noise were recorded from the cochlear nucleus of anesthetized chinchillas. Rippled noises had a fixed delay of 4 ms, and spectral depth was varied by attenuating the delayed version of the noise. Temporal discharge patterns were analyzed using neural autocorrelograms, and responses to rippled noises were compared to wideband noise responses. Chopper units with best frequencies in the range of the first to second harmonics of the rippled noise showed large differences in discharge patterns between rippled noise and wideband noise responses, but chopper units with best frequencies centered at higher harmonics did not show large differences. Consequently, the Chopper group of units showed no evidence of a neural representation of the dominance region. Primarylike units did show a neural representation of dominance that is related to behavioral performance. For Primarylike units with best frequencies around the third to fifth harmonics of the rippled noise, large differences in discharge patterns between rippled noise and wideband noise responses were observed. The results suggest that bushy cells play an important role in processing pitch-related information and should be included as important elements in neural models of periodicity processing.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2981637]

PACS number(s): 43.64.Qh, 43.66.Hg, 43.80.Lb [BLM] Pages: 3038–3052

I. INTRODUCTION

Pitch is a fundamental perception associated with a variety of complex periodic sounds. In music perception, changes in pitch form the basis of melody recognition. In speech perception, changes in pitch form the basis of prosody in languages such as English, but for tonal languages like Mandarin, changes in pitch of a monosyllabic word result in a change in the meaning of the word. Pitch also plays a role in the ability of human listeners to segregate sounds from many simultaneous sources (i.e., auditory scene analysis).

A variety of psychological attributes of pitch have been described in human listeners [see Plack *et al.* (2005)], and a principal attribute is the existence of a spectral dominance region of pitch. In general, for a sound comprised of successive harmonic tones, the frequencies most effective in evoking the pitch perception are found in the region of the third to fifth harmonics (Plomp, 1967; Ritsma, 1967). The existence of the dominance region of pitch is not limited to harmonic complex tones but has also been demonstrated in human listeners using pitch discrimination tasks with bandpass filtered rippled noises (Bilsen and Ritsma, 1970; Yost and Hill, 1978; Yost, 1982; Leek and Summers, 2001). In addition, the existence of a spectral dominance region does not appear to be a perceptual attribute that is the result of a uniquely human neural mechanism. Behavioral studies using operant conditioning in chinchillas have shown the existence of a dominance region for the discrimination of rippled noise from flat-spectrum wideband noise (WBN) (Shofner and Yost, 1997). The discrimination of rippled noise from WBN is known as coloration discrimination, and the location of dominance in the chinchilla for coloration discrimination also corresponds to the spectral region around the third to fifth harmonics.

Physiological correlates of spectral dominance for tone complexes have been described for the frequency following response (FFR) recorded from humans (Greenberg *et al.*, 1987). The FFR is a scalp-recorded, auditory evoked potential that represents the synchronous or phase-locked activity of auditory brainstem neurons to periodic sounds (Marsh *et al.*, 1972). The amplitude of the FFR recorded in response to two successive harmonics of a fundamental frequency was largest when the third and fourth harmonics were presented (Greenberg *et al.*, 1987). Further evidence of the role of neural synchrony in the representation of the dominance region can be found from the responses of single auditory nerve fibers (Cariani and Delgutte, 1996). Interspike intervals (ISIs) of auditory nerve fibers were measured in response to two different harmonic tone complexes presented simultaneously. One complex was made of the third to fifth harmonics of a fundamental frequency, whereas the other complex was made of the 6th–12th harmonics of a different fundamental frequency. For fundamental frequencies of 160 Hz and above, the predominant ISI corresponded to the period of the fundamental frequency associated with third to fifth harmonics (Cariani and Delgutte, 1996). The findings of these two studies suggest that dominance is related to neural

a)Electronic mail: wshofner@indiana.edu

 © 2008 Acoustical Society of America

synchrony and is encoded as early as the auditory nerve. Thus, a neural representation of dominance would be expected to exist in the cochlear nucleus, but do all cochlear nucleus unit types show a representation of dominance in their temporal discharge patterns?

The cochlear nucleus receives its ascending input from the auditory nerve and is the first nucleus along the central auditory pathway where neural processing can occur. The nucleus is divided into three major subdivisions: anteroventral cochlear nucleus, posteroventral cochlear nucleus, and dorsal cochlear nucleus (Osen, 1969; Morest *et al.*, 1990). Each of these subdivisions contains several distinct principal cell subsystems that differ in their neuronal morphologies, innervation, and physiological response properties. The principal cells of the anteroventral division are the bushy cells and the stellate or multipolar cells; primarylike responses are associated with bushy cells (Rhode *et al.*, 1983; Rouiller and Ryugo, 1984; Smith and Rhode, 1987; Smith *et al.*, 1991, 1993; Ostapoff *et al.*, 1994), whereas chopper responses are associated with stellate or multipolar cells (Rhode *et al.*, 1983; Rouiller and Ryugo, 1984; Smith and Rhode, 1989; Ostapoff *et al.*, 1994). Multipolar cells receive auditory nerve input in the form of small bouton terminals (Redd *et al.*, 2002), whereas bushy cells receive large end bulb or calyx synaptic terminals from auditory nerve fibers (Cant and Morest, 1979; Lorente de No, 1981; Ryugo and Fekete, 1982; Sento and Ryugo, 1989; Ryugo and Sento, 1991). The depolarization that occurs in end bulb synapses can be recorded extracellularly as prepotentials (Pfeiffer, 1966; Guinan and Li, 1990). Although both unit types can give synchronous responses to rippled noise stimuli that are correlated with the pitch perceived in human listeners (Shofner, 1991, 1999; Wiegrebe and Winter, 2001; Winter *et al.*, 2001; Verhey and Winter, 2006; Sayles and Winter, 2007), the specific functional roles that primarylike and chopper units play in periodicity processing as related to pitch remain unclear. Chopper units show stronger phase-locked responses than primarylike units to sinusoidal-amplitude modulated tones (Frisina *et al.*, 1990; Rhode and Greenberg, 1994; Shofner *et al.*, 1996) as well as to the fundamental frequency of synthetic vowels (Wang and Sachs, 1994). Consequently, chopper units have been thought to play an important role in periodicity processing, and most attempts to model periodicity processing utilize either chopper units exclusively (Meddis and O'Mard, 2006) or as one component of the neural circuit (Borst *et al.*, 2004; Dicke *et al.*, 2007). Although none of these models includes contributions from primarylike units, Nelson and Carney (2004) developed an alternative model based exclusively on the contributions of primarylike units rather than chopper units.

The present study seeks to extend the previously reported physiological findings into the cochlear nucleus of chinchillas using rippled noise stimuli. The approach of the present experiment is not to study the responses of a single unit to a variety of pitch-evoking stimuli but rather to examine the neural representation across a population of units to a limited set of stimuli. This approach is similar to that previously taken to study the encoding of steady-state vowels [e.g., see Blackburn and Sachs (1990) and Winter and Palmer (1990b)]. Stimuli were chosen to allow comparisons between single-unit responses and behavioral performance obtained previously in the chinchilla (Shofner and Yost, 1997). The present report describes a neural correlate of the dominance region in the temporal discharge of primarylike units but not of chopper units. The results argue that in addition to chopper units (stellate cells), primarylike units (bushy cells) play an important role in processing pitch-related information and should also be included in neural models of periodicity processing.

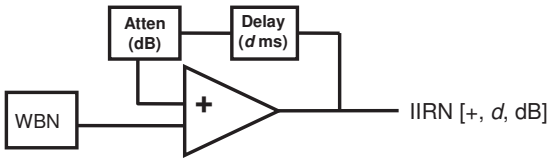


FIG. 1. Schematic diagram shows the positive feedback circuit used for generating infinitely iterated rippled noise (IIRN [+, *d*, dB]) from a wideband noise (WBN) input, where + indicates positive gain (i.e., added), *d* is the delay, and dB is the delayed noise attenuation.

A. Properties of iterated rippled noises

Rippled noises have been used to study a wide range of auditory phenomena, including masking and frequency selectivity (Houtgast, 1977; Pick, 1980; Niemiec *et al.*, 1992) and echo processing (Bassett and Eastmond, 1964), and as a general stimulus for complex sound processing (Schreiner and Calhoun, 1994; Shamma *et al.*, 1995; Amagai *et al.*, 1999). Moreover, rippled noises have become an important class of stimuli for studying the perception of pitch and are important stimuli for testing models of pitch perception (Meddis and Hewitt, 1991; Cohen *et al.*, 1995; Yost *et al.*, 1996).

Rippled noises are generated when a WBN is delayed [*d* (ms)] and the delayed noise is added to the undelayed version of the noise. If the rippled noise is delayed again and added to the original WBN, the output is an iterated rippled noise of two iterations (i.e., two iterations of delay and add). An infinitely iterated rippled noise (IIRN) can be achieved using a positive feedback loop in the delay and add process (Fig. 1). For convenience, IIRNs will be referred to using the following notation: IIRN [+, *d*, dB], where + indicates that the delayed noise is added, *d* is the delay in ms, and dB is the attenuation of the delayed noise relative to the original undelayed noise (see Fig. 1). Unlike complex tones, which are characterized by line spectra, rippled noises are characterized by continuous spectra, and sample spectra based on the WBN input [Fig. 2(A)] to the positive feedback network described in Fig. 1 are shown in Figs. 2(B) and 2(C) for a 4 ms delay. When the delayed noise is added to the original noise, the spectrum of IIRN has peaks at integer multiples of $1/d$; this is the harmonic condition of IIRN. When the delayed noise attenuation is decreased from −3 to −1 dB, there is an increase in the peak-to-valley ratio of the spectral peaks.

The temporal properties of rippled noises are best described by autocorrelation functions, and sample autocorrelation functions based on a WBN noise input [Fig. 2(D)] to the positive feedback network described above are also

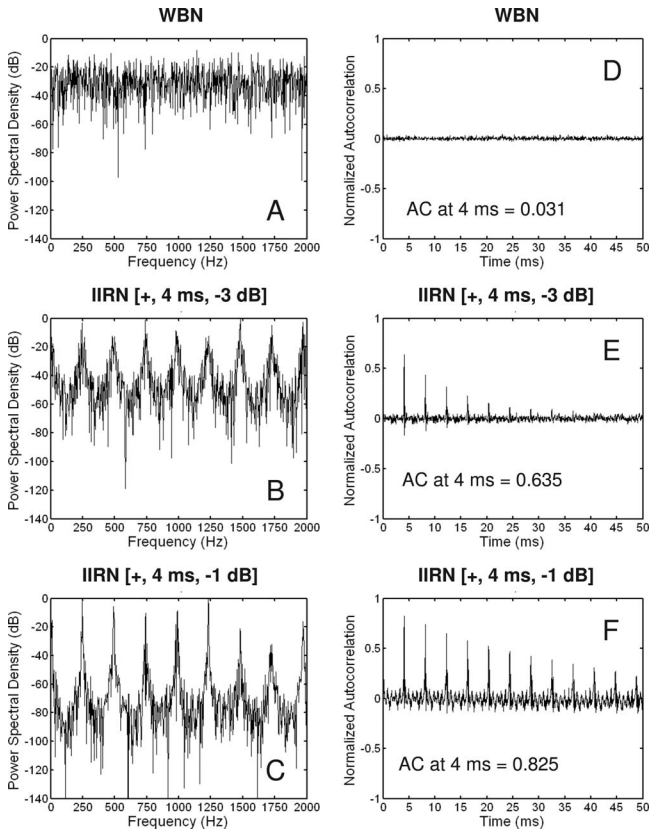


FIG. 2. [(A)–(C)] Sample spectra are shown for samples of the WBN input (A) and IIRNs with a fixed delay (d) of 4 ms and delayed noise attenuations of −3 dB (B) and −1 dB (C). The bandwidths of the noises were 10 kHz, but for purposes of clarity, only frequencies up to 2000 Hz are shown. Note for the IIRN spectra, the peaks occur at integer multiples of 250 Hz, i.e., $1/d$. The spectral depth of the ripple increases as the amount of delayed noise attenuation decreases. [(D)–(F)] Sample autocorrelation functions of the waveform are shown for the same samples of WBN and IIRNs. The height of the autocorrelation at 4 ms (AC at 4 ms) is given for each stimulus. Note that both the height and number of the peaks increase as the delayed noise attenuation decreases.

shown in Figs. 2(E) and 2(F). When the delayed noise is added to the original noise, the waveform autocorrelation functions of IIRN stimuli show positive correlations at time lags corresponding to integer multiples of d. The heights of these peaks increase as the amount of delayed noise attenuation decreases from −3 to −1 dB. Unlike complex tones, which are periodic, rippled noises possess a regularity in their temporal structure but are not truly periodic. Rather, the correlations in the IIRN autocorrelation functions reflect a part of the waveform that repeats itself throughout the stimulus waveform but does not repeat itself in a periodic manner. Thus, rippled noises have been described as pseudoperiodic or quasiperiodic waveforms, and the height of the first peak in the autocorrelation function can be taken as a measure of the strength of the regularity [Figs. 2(E) and 2(F)].

Rippled noises evoke pitches in human listeners that are the same as the pitches evoked by the many other classes of stimuli that produce complex pitches. For iterated rippled noise stimuli, the pitch is determined by d. In general, the pitch of iterated rippled noise with positive gain (i.e., added) is at a frequency corresponding to $1/d$ and corresponds to the time lag of the first positive correlation in the waveform autocorrelation function (Yost, 1996a). The pitch strength of rippled noises can be made to vary along a continuum from a flat-spectrum noise (having no real pitch) to a periodic pulse train (having a very strong pitch). The pitch strength increases as either the number of iterations increases (Yost, 1996b) or the amount of delayed noise attenuation decreases (Shofner and Selas, 2002).

II. METHODS

A. General procedures

The experiments were carried out at Loyola University Chicago and Indiana University, and the procedures have been reviewed and approved by the Institutional Animal Care and Use Committees of both Loyola University Chicago and Indiana University. Adult chinchillas weighing 500–800 g were anesthetized with intraperitoneal injections of sodium pentobarbital. For recordings obtained at Loyola University, animals were initially injected with a dose of 70 mg/kg; for recordings obtained at Indiana University, animals were initially injected with a dose of 60 mg/kg and given a subcutaneous injection of Rimadyl (4 mg/kg) as additional analgesic. Supplemental injections of pentobarbital were given to maintain areflexia. Body temperature was maintained at around 38 °C with a dc heating pad.

A tracheotomy was performed, and the external auditory canals were exposed and transected. Animals were placed in a modified headholder (Kopf Model 900), and the left bulla was exposed and opened in order to expose the cerebellum as originally described by Frisina *et al.* (1982) for the gerbil. An opening was made in the temporal bone to expose the cerebellum, and electrodes were advanced through the cerebellum into the underlying cochlear nucleus using a hydraulic microdrive system (Kopf 650). The advantage of this transbullar approach to the cochlear nucleus is that the recording of a single-unit activity can be very stable; it is not uncommon to hold units for 1–2 h. Any effect opening the bulla may have on the middle ear transfer function [e.g., see Ruggero *et al.* (1990)], and thus on neural thresholds, should be equal across all unit types in the cochlear nucleus.

B. Neurophysiological recording and stimuli

Animals were placed in a double-walled sound attenuating chamber during neurophysiological recording of single-unit responses. Commercially purchased tungsten microelectrodes were used to record single-unit activities. Data acquisition and stimulus presentation were under the control of a Gateway computer system and Tucker-Davis Technologies System II modules. Extracellular action potentials were amplified, filtered, and passed through a spike discriminator. The times of occurrences of the pulses were determined online relative to the onset of the stimulus with 1 μs resolution. Data from the neural activity was only recorded from clearly isolated single-units as determined by monitoring the triggered spike waveform online. Single-units were not isolated from multiunit clusters of spikes. Acoustic stimuli were presented to the left ear through a Sennheiser HD 414 SL earphone enclosed in a brass housing that also held a calibration

microphone. Search stimuli were 100 ms bursts of either WBN or tones at the best frequency (BF) of the background neural activity.

When a unit was isolated, its BF was first determined using audiovisual cues, and data were collected for unit classification. The classification of unit types was based on poststimulus time (PST) histograms, ISI histograms, regularity histograms (Young *et al.*, 1988), and the presence or absence of a prepotential. At least 1000 amplified spikes were digitized through an analog-digital (A/D) channel and were averaged online in order to determine if prepotentials were present in the averaged spike waveform. BF tones were generated by the computer and presented through a 16 bit digital-analog (A/D) converter. The sampling rate of the A/D and the conversion rate of the D/A were 50 kHz. BF tones of 200 ms duration presented once every 1000 ms with rise/fall times of 10 ms were used to generate rate-level functions over a 100 dB range in 1 dB steps. One stimulus was presented at each level, and the rate-level function was smoothed using a five-bin triangular moving window average. The lowest stimulus level presented was well below the audiovisual threshold, and the smoothed firing rate produced at the ten lowest levels was used to estimate the mean and standard deviation of the spontaneous discharge rate. The threshold was then defined as the level that first evoked an increase in discharge rate greater than 2 standard deviations above the spontaneous discharge rate, provided that the next three levels were also greater than 2 standard deviations above the spontaneous rate. PST histograms were then generated for 250 presentations of a 50 ms BF tone with 2 ms rise/fall times presented once every 250 ms at levels of 20–40 dB above the threshold. If a PST histogram showed strong phase-locking such that a characteristic discharge pattern was obscured, then BF tones were also presented with random starting phases in an attempt to ascertain a characteristic discharge pattern. Spike times were also used to generate ISI histograms and regularity histograms. Spikes that occurred during the initial 20 ms were not used for generating ISI histograms in order to ensure that ISI histograms characterized the steady-state portion of the response. The mean ISI and standard deviation were determined from pairs of spikes in which the times of occurrence were 20–50 ms. The coefficient of variation (CV) was computed as the ratio of the standard deviation to the mean.

After the data were collected for unit classification, the responses to WBN and IIRN were studied. Sets of WBN and IIRN stimuli were generated using the same equipment and parameter settings that were previously used to measure discrimination performance in chinchillas (Shofner and Yost, 1997). For WBN and each IIRN, 5 s of the waveform was sampled at 50 kHz; the waveform amplitudes were then adjusted so that all WBN and IIRN stimuli had equal root-mean-squared amplitudes. The 5 s samples of each noise were stored on a disk as stimulus files. For a stimulus presentation, WBN and IIRN signals were presented through a 16 bit D/A converter (TDT DA3-2 module) at a conversion rate of 50 kHz. During a stimulus presentation for a particular noise signal, a separate 500 ms sequence was randomly obtained from the 5 s sample for each stimulus presentation. Consequently, the waveform of the noise was not the same for each presentation.

Rate-level functions were first generated over a 100 dB range in 1 dB steps for 500 ms WBN bursts with rise/fall times of 10 ms presented once every 1500 ms. One noise burst was presented at each level, and the rate-level function was smoothed using a five-bin triangular moving window average. The threshold for the WBN was estimated as defined above for BF tone rate-level functions. After the WBN rate-level function had been obtained, the responses of the isolated unit were studied for WBN and IIRN stimuli. A total of 100 separate samples of WBN or IIRN were presented, having durations of 500 ms and rise/fall times of 10 ms; bursts were presented once every 1000 ms. These were the same stimulus parameters used to measure behavioral performance (Shofner and Yost, 1997). Typically, data were first obtained for WBN and then for a series of IIRN stimuli with a fixed delay of 4 ms and delayed noise attenuations of −1 to −4 dB. The overall levels of the WBN and IIRN stimuli were presented at 20 dB above the threshold as determined from the WBN rate-level function.

Because multiple electrode tracks were made in a typical experiment, the histological location of units was not determined. However, recovered lesions and electrode tracks that were verified histologically from past experiments (Shofner, 1991; Shofner *et al.*, 1996) using the same surgical approach and angle of penetration were limited to ventral cochlear nucleus (VCN). In addition, the large proportion of units with prepotentials in the present sample strongly suggests many electrode penetrations passed through the anteroventral cochlear nucleus (AVCN).

C. Data analysis

Temporal responses of units to WBN and IIRN were analyzed as autocorrelograms (i.e., all-order ISI histograms). The autocorrelogram displays the average discharge pattern following a given action potential. Autocorrelograms were generated for action potentials occurring between 20 and 500 ms of the stimulus; spikes that occurred during the initial 20 ms were not used in order to ensure that autocorrelograms characterized the steady-state portion of the response. The ordinates of the autocorrelograms were first scaled in terms of firing rate as described by Abeles (1982), and were then scaled in terms of normalized firing rate (λ_τ) as

$$\lambda_\tau = \frac{R_\tau - R_{\text{ave}}}{R_{\text{ave}}}, \tag{1}$$

where R_τ is the firing rate at a time lag of τ ms and R_{ave} is the average firing rate. Normalized rate is a unitless measure, and a normalized rate of 0 represents the average firing rate. The difference in temporal response patterns between responses to IIRNs and WBN was quantified by using the fourth moment of the normalized autocorrelogram (see Sec. IV). The fourth moment can be used as a measure of the instantaneous fluctuations in power in a waveform (Hartmann, 1998), and the normalized autocorrelogram is treated

as a waveform. The average fourth moment of the normalized autocorrelogram was computed as

$$\overline{\lambda^4} = \frac{\Sigma(\lambda_\tau - \lambda_{ave})^4}{N}, \quad (2)$$

where N corresponds to a 50 ms window (i.e., 500 bins that are each 100 μs wide). Because λ_{ave} is equal to 0 in the normalized autocorrelogram, the average fourth moment becomes

$$\overline{\lambda^4} = \frac{\Sigma(\lambda_\tau)^4}{N}. \quad (3)$$

The ratio of the averaged fourth moments for IIRN and WBN responses (i.e., relative fourth moment) was expressed in decibel and given as

$$\text{dB} = 10 \log \frac{\overline{\lambda^4}_{\text{IIRN}}}{\overline{\lambda^4}_{\text{WBN}}}. \quad (4)$$

A large decibel indicates greater deviations from a normalized rate of 0 in the IIRN response relative to the deviations around 0 in the WBN response.

III. RESULTS

A. Unit classification

The results of the present study are based on the responses of 118 single-units recorded from the cochlear nucleus of adult chinchillas. Units were classified on the basis of their discharge characteristics generated for BF tone bursts at 20–40 dB above the threshold, as described in Sec. II. For purposes of clarity, the use of the word "Primarylike" will be used to refer specifically to the group of prepotential and primarylike units obtained in the present study, whereas the word "Chopper" will be used specifically in reference to the sample of chopper and regular units obtained in this study.

A total of 48 units were classified as Primarylike; in response to BF tones, these units show an initial increase in firing rate that adapts to a relatively constant steady-state discharge [Fig. 3(A)]. Of the units classified as Primarylike, 26 units were observed to have a prepotential in the averaged spike waveform (see the inset in Fig. 3). The analysis of the discharge pattern of this primarylike with a prepotential unit based on regularity histograms (Young *et al.*, 1988) and ISI histograms indicates an irregular discharge of spikes during the tone [Figs. 3(B) and 3(C)]. The CV calculated from the ISI histogram shown in Fig. 3(C) is 0.613, and for BF tones at 20–40 dB above the threshold for the sample of prepotenital units, the CVs ranged from 0.5–0.8 with a mean of 0.67.

A total of 64 units were classified as Chopper units. In this sample, 34 units showed a series of peaks in the PST histogram; this observed chopping pattern is the result of a precisely timed onset spike followed by a regular discharge of spikes [Fig. 3(D)]. The regularity histogram [Fig. 3(E)] and ISI histogram [Fig. 3(F)] indicate a regular discharge pattern; the CV calculated from the ISI histogram in Fig. 3(F) is 0.366. The mean CV calculated from the ISI histograms in response to BF tones at 20–40 dB above the threshold for the sample of units with chopper PST histograms was 0.41. The remaining 30 units in this group did not show the characteristic chopping discharge pattern in the PST histograms but had CVs of ISI histograms that were smaller than those of prepotential units, indicating a more regular discharge that is characteristic of chopper units. The mean CV calculated from the ISI histograms in response to BF tones at 20–40 dB above the threshold for the sample of regular units was 0.43.

A single-factor analysis of variance (ANOVA) was carried out on the CV calculated from the ISI histograms for the prepotential units, regular units, and chopper units. Because primarylike PST histograms can be recorded from cell types other than bushy cells, while the presence of a prepotential is a signature characteristic of a bushy cell, the ANOVA only included prepotential units as a quantitative estimate of the irregular discharge of bushy cells. A significant effect across groups was found for the mean CVs ($F=78.43$, $P<0.0005$). Paired comparisons were made using Tukey's test. At $\alpha=0.001$, there was a significant difference between the mean CVs of the prepotential and regular units and between the mean CVs of the prepotential and chopper units. The difference between mean CVs of the chopper units and regular units was not significant ($P>0.05$). Consequently, regular and chopper units were combined to form the Chopper group. These regular units are considered to be Chopper units that do not have a precisely timed onset spike and thus do not show the characteristic chopping in the PST histogram [see Winter and Palmer (1990a); and Shofner (1999)].

Units having low BFs (<500 Hz) typically show strong phase-locking to BF tones, thus making it difficult to distinguish a characteristic discharge pattern. Low BF units were classified as Primarylike if the CV was greater than 0.6 and were classified as regular if the CV was less than 0.5. These low BF regular units were included in the Chopper category. A total of six units had CVs that were 0.5-0.6. Thus, it was difficult to determine on this basis if they were primarylike or regular. These six units were classified as phase locked.

B. Responses to infinitely iterated rippled noise

The temporal discharge patterns obtained in response to WBN and to IIRNs with a fixed delay of 4 ms and delayed noise attenuations of −1, −2, and −3 dB are shown in Figs. 4(A)–4(C) for the primarylike unit with a prepotential illustrated in Fig. 3. Discharge patterns are displayed as normalized neural autocorrelograms, which show the average firing pattern of the unit following an action potential that occurs at a time lag of 0. A normalized firing rate of 0 is equal to the average firing rate of the unit (see Sec. II). For comparison, the temporal discharge patterns in response to WBN are also shown. The normalized autocorrelogram obtained in response to WBN is relatively flat, indicating a random nonperiodic discharge pattern. In contrast, the normalized autocorrelograms obtained in response to IIRNs show a series of peaks and nulls. The peaks occur at 4 ms and integer multiples of 4 ms, indicating that there is a periodicity to the discharge pattern of the unit that is related to the IIRN delay.

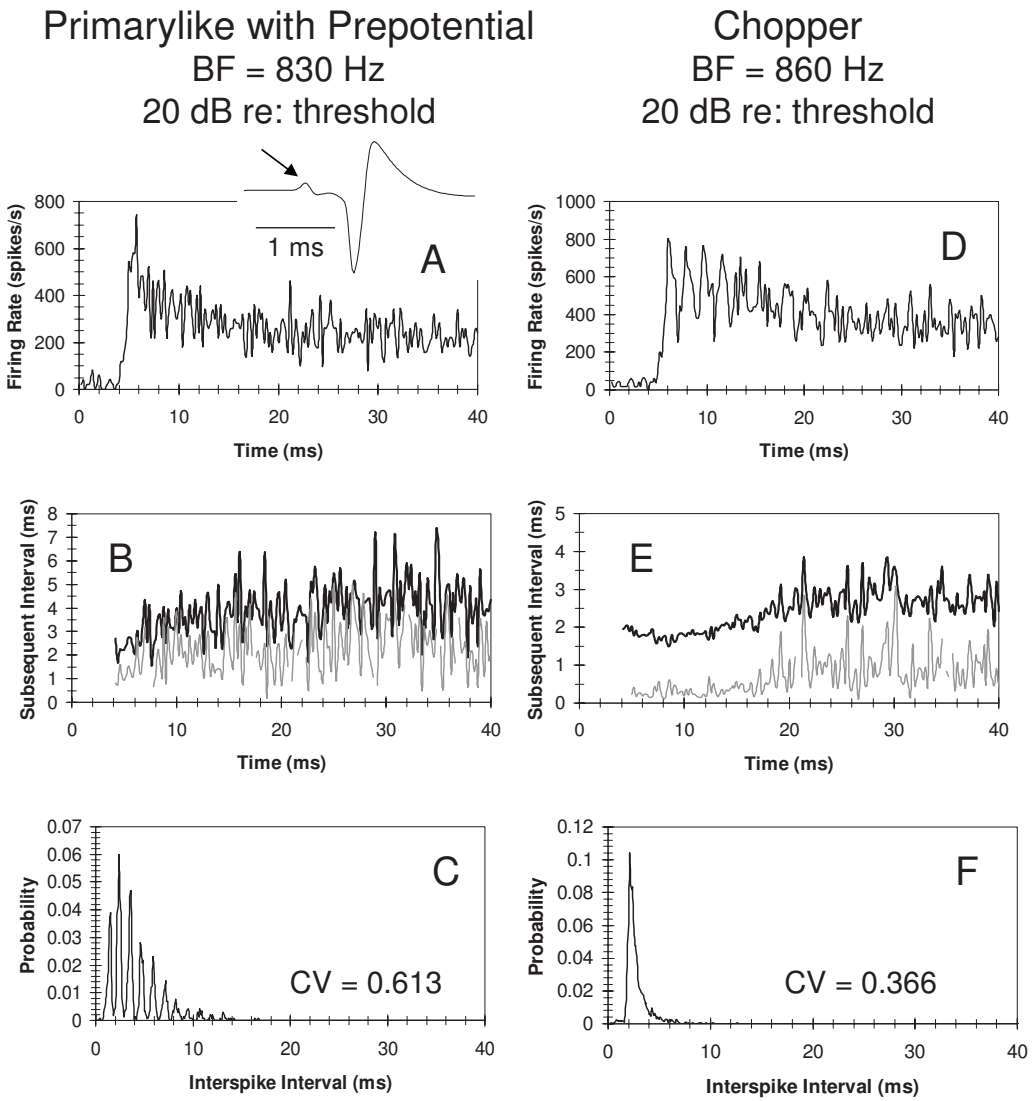


FIG. 3. Discharge properties of a primarylike unit and a chopper unit with similar best frequencies (BFs) are illustrated. Responses were obtained from BF tones at 20 dB above the threshold. The primarylike unit in this example exhibited a prepotential in the averaged spike waveform (see arrow in the inset). (A) PST histogram obtained from the primarylike unit. (B) Regularity histogram obtained for the primarylike unit shows mean subsequent ISI (black line) as a function of time through the PST histogram and the standard deviation of the subsequent interval (gray line). (C) ISI histogram obtained from the primarylike unit. The CV calculated from ISI histogram was 0.613. The interval analysis shown in (B) and C indicates an irregular discharge pattern. (D) PST histogram obtained from a chopper unit. Note the peaks and valleys in the onset portion of the histogram. (E) Regularity histogram obtained for the chopper unit. Black and gray lines same as in (B). (F) ISI histogram obtained from the chopper unit. The CV calculated from the ISI histogram was 0.366. The interval analysis shown in (E) and (F) indicates a more regular discharge pattern for the chopper unit than for the primarylike unit.

Notice that the number of peaks and the heights of the peaks both decrease as the delayed noise attenuation increases from −1 to −3 dB [blue lines in Figs. 4(A)–4(C)]. That is, as the amount of delayed noise attenuation increases, the periodicity of the discharge pattern decreases, becoming more like that of WBN (see the red lines in Fig. 4). Thus, for the autocorrelograms illustrated in Figs. 4(A)–4(C), the deviations around a normalized firing rate of 0 are small for WBN, but are large for IIRNs. These deviations are a reflection of the periodic discharge of the unit in response to IIRN with larger deviations, indicating a stronger periodicity of discharge or a more synchronous discharge.

The normalized autocorrelogram can be treated as a waveform, and for any waveform, the fourth moment can be used as a measure of the instantaneous fluctuations in power (Hartmann, 1998). The difference in discharge patterns between WBN and IIRN responses was quantified by computing the average fourth moment of the normalized autocorrelogram over a 50 ms time window (see Sec. II). The relative fourth moment expressed as a decibel was then computed (see Sec. II) in order to compare the response to an IIRN to the WBN response. Because chinchillas discriminated IIRN from WBN in the behavioral task (Shofner and Yost, 1997), it would be expected that the underlying neural processing would involve a comparison of responses between IIRN and WBN. For the autocorrelograms of the primarylike unit shown in Fig. 4, the relative fourth moments are 11.56, 5.02, and 1.36 dB for IIRNs with delayed noise attenuations of −1, −2, and −3 dB, respectively. Larger relative fourth moments indicate a greater difference in the discharge patterns between IIRN and WBN responses. For comparison, Figs. 4(D)–4(F) also show the autocorrelograms obtained in re-

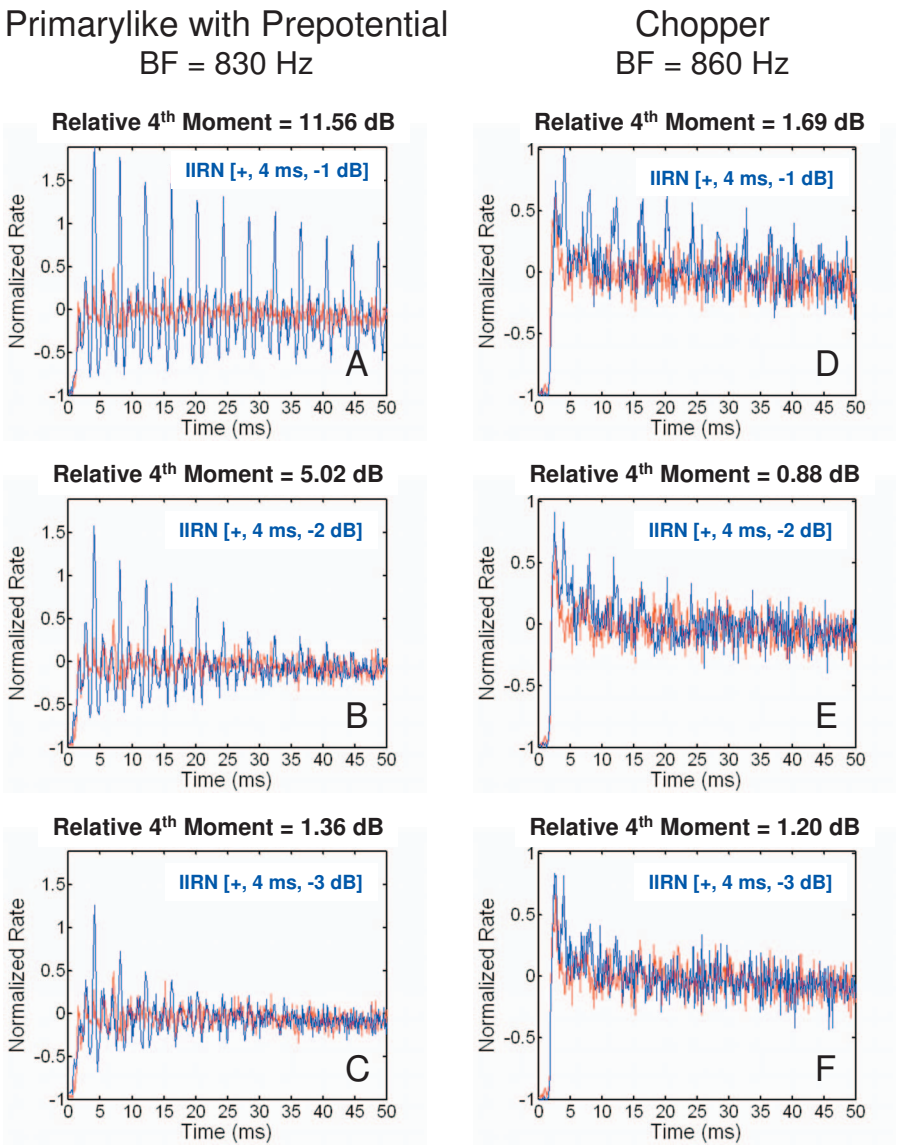


FIG. 4. Normalized neural autocorrelograms for the primarylike unit [(A)–(C)] and chopper unit [(D)–(F)] characterized in Fig. 3 are illustrated. Blue lines show the autocorrelograms obtained in response to IIRNs; red lines show the autocorrelograms obtained in response to WBN. The delayed noise attenuation are −1 dB [(A) and (D)], −2 dB [(B) and (E)], and −3 dB [(C) and (F)]. The relative fourth moment as defined in the text is given for each panel.

sponse to WBN and IIRNs for a chopper unit having a similar BF. Note that in this example, the deviations around a normalized rate of 0 are not as pronounced as those for the primarylike unit, and consequently, the values of the relative fourth moments are smaller. For the autocorrelograms of the chopper unit shown in Figs. 4(D)–4(F), the relative fourth moments are 1.69, 0.88, and 1.20 dB for IIRNs with delayed noise attenuations of −1, −2, and −3 dB, respectively. These smaller values of the relative fourth moment reflect less of a difference between the discharge patterns for IIRN and WBN responses for this chopper unit. Note also that in response to the IIRN with the larger delayed noise attenuation of −3 dB, there is a clear periodicity to the discharge pattern for the primarylike unit but not for the chopper unit [Figs. 4(C) and 4(F)].

Neural autocorrelograms for a low BF primarylike unit [Figs. 5(A)–5(C)] and a low BF regular unit [Fig. 5(D)–5(F)] are similar to the responses previously shown in Fig. 4 for primarylike and chopper units in that large differences in discharge patterns are observed between IIRN [+, 4 ms, −1 dB] and WBN. That is, a comparison of the responses in Figs. 4(A), 4(D), 5(A), and 5(D) shows periodicities in the responses for all units related to the 4 ms delay. Also, for both the low BF primarylike and regular units, the differences in discharge patterns between IIRN and WBN decrease as the delayed noise attenuation increases, as shown by the decreases in the relative fourth moments (Fig. 5). Also note that the difference in discharge patterns between IIRN [+, 4 ms, −4 dB] and WBN is larger for the regular unit [Fig. 5(F)] than for the primarylike unit [Fig. 5(C)]. This example contrasts with the previous example shown in Fig. 4 in which the response to the IIRN was larger for the primarylike unit [Fig. 4(C)] than for the chopper unit [Fig. 4(F)] at the larger delayed noise attenuation of −3 dB.

Scatter plots of the relative fourth moment as a function of BF are shown in Fig. 6 for individual Primarylike units in response to 4 ms delayed IIRN having delayed noise attenuations of −1, −2, −3, and −4 dB. The neural profiles of the relative fourth moment across BFs are shown by the black squares and heavy solid lines. These profiles are the relative fourth moment averaged in one-octave widebands that are centered at the harmonic frequencies of the IIRN (i.e., at 250, 500, 750, 1000 Hz, etc.) and are estimates of the population responses across BFs. Note that these are moving window

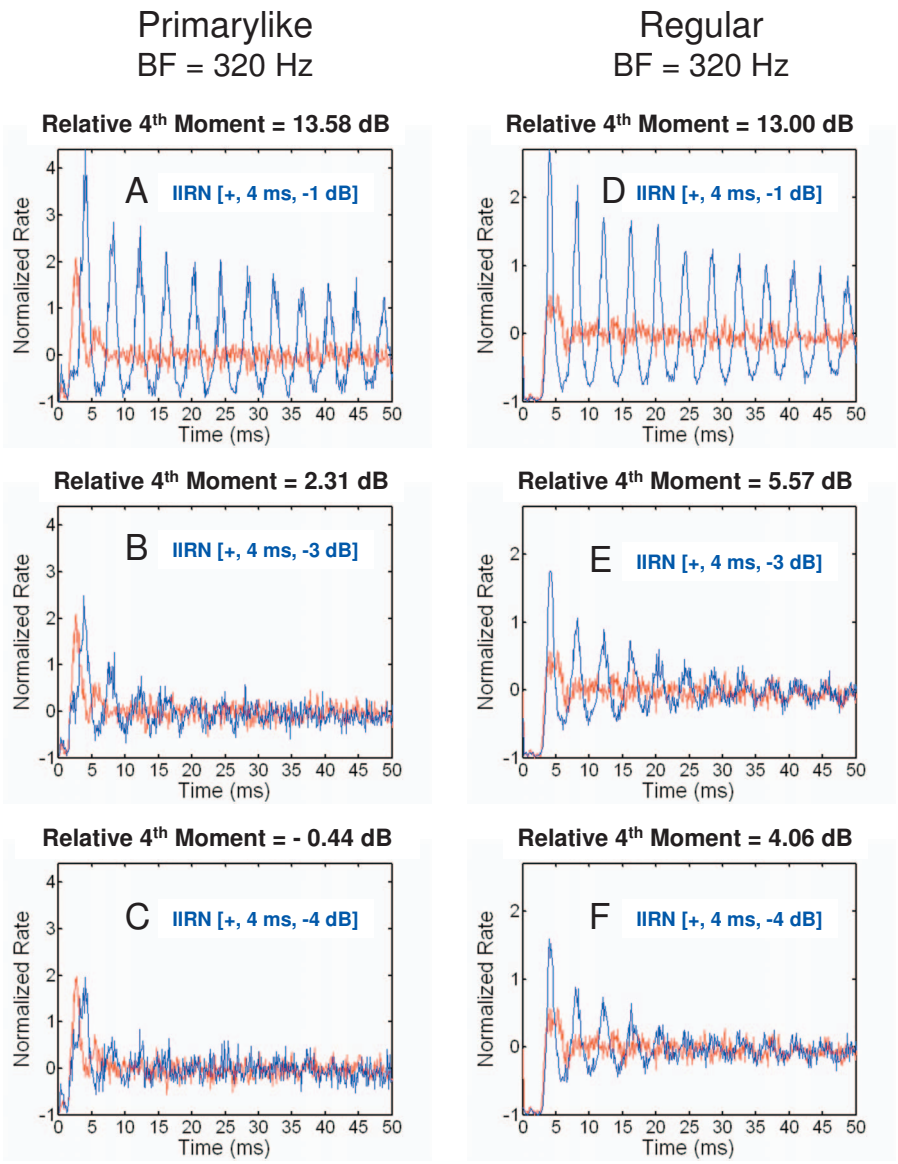


FIG. 5. Normalized neural autocorrelograms for a primarylike unit [(A)–(C)] and a regular unit [(D)–(F)] having BFs of 320 Hz are illustrated. Blue lines show the autocorrelograms obtained in response to IIRNs; red lines show the autocorrelograms obtained in response to WBN. The delayed noise attenuation is −1 dB [(A) and (D)], −3 dB [(B) and (E)], and −4 dB [(C) and (F)]. The relative fourth moment as defined in the text is given for each panel.

averages, and this averaging within one-octave bands is similar to the one-octave bandpass filtering used to collect behavioral data from chinchillas (Shofner and Yost, 1997). A single-factor ANOVA was carried out on the relative fourth moment for Primarylike units having BFs less than or equal to 1250 Hz (i.e., less than or equal to the fifth harmonic of the IIRN), with the four delayed noise attenuations as the factor. A significant effect of delayed noise attenuation was found on the mean relative fourth moment for the Primarylike group ($F=10.52$; $P<0.0005$). Scatter plots of the relative fourth moment as a function of BF for individual Chopper units are shown in Fig. 7. A single-factor ANOVA was carried out on the relative fourth moment for Chopper units having BFs less than or equal to 1250 Hz for the four delayed noise attenuations. A significant effect of delayed noise attenuation was found on the mean relative fourth moment for the Chopper group ($F=7.67$; $P<0.0005$).

The neural profiles of the relative fourth moment averaged in one-octave widebands centered at the harmonic frequencies of the IIRN are shown by the black squares and heavy solid lines in Figs. 6 and 7. Although it is not clear from the scale used in these figures, the shape of the neural profiles appears to be different for the Primarylike and Chopper groups. Figure 8 compares the neural profiles for the two groups of units on the same ordinate scale in order to highlight differences in the shapes of the neural profiles. The harmonic number in Fig. 8 indicates the locations of the spectral peaks for the 4 ms IIRN. The profiles obtained for the Primarylike group appear to have more of a lowpass shape for delayed noise attenuations of −1 and −2 dB but more of a bandpass shape for delayed noise attenuations of −3 and −4 dB. In contrast, the profiles obtained for the Chopper group have a definite lowpass shape at all delayed noise attenuations, and the cutoff frequency appears to occur at a lower harmonic than the cutoff frequency for the Primarylike group (Fig. 8). The averaged relative fourth moments for the first and second harmonics are similar for Primarylike and Chopper groups, but are larger at the third to fifth harmonics for Primarylike units than for Chopper units (Fig. 8). The neural profiles shown in Fig. 8 are moving window averages, and as such, there is some overlap in the relative fourth moments between successive harmonics. Conse-

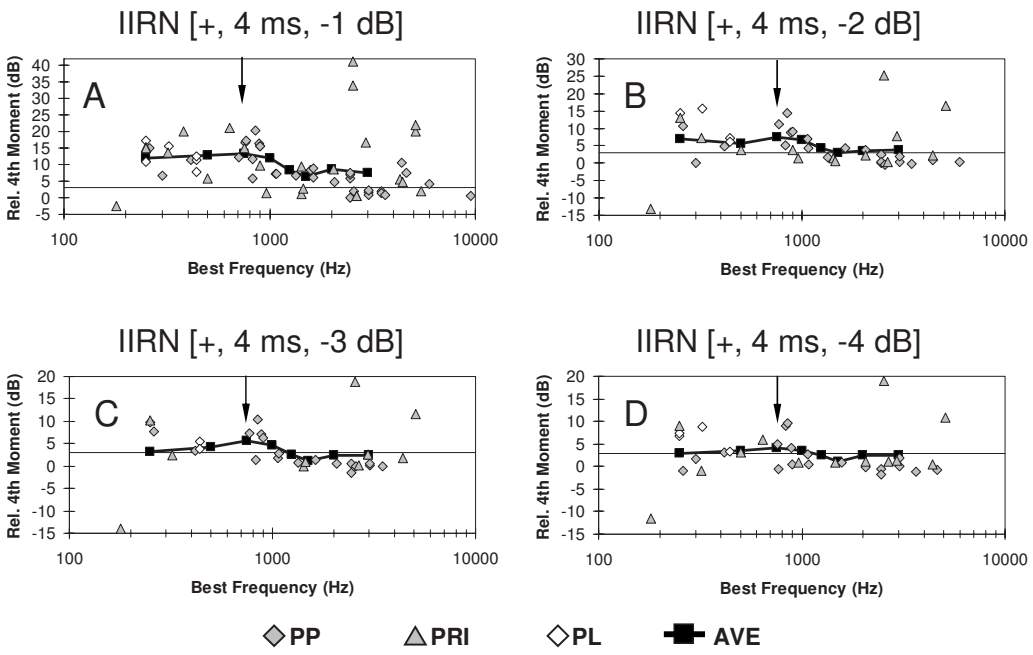


FIG. 6. Scatter plots illustrating the relative fourth moment as a function of BF for the Primarylike group for delayed noise attenuations of −1 dB (A), −2 dB (B), −3 dB (C), and −4 dB (D). Gray diamonds show units having a prepotential (PP); gray triangles show units with primarylike PST histograms but without a prepotential (PRI). The open diamonds illustrate the relative fourth moments obtained from low BF phase-locked (PL) units that could not be characterized as either Primarylike or Chopper. The black squares and solid black line show the relative fourth moment averaged in one-octave bands (ave), centered at harmonic frequencies of the rippled noise (i.e., 250, 500, 750, 1000 Hz, etc). The one-octave bandpass averaging is similar to the one-octave bandpass filtering used to collect behavioral data (Shofner and Yost, 1997). For reference, the downward pointing arrow indicates the frequency of the third harmonic. Note that the ordinate changes scales in (A)–(D).

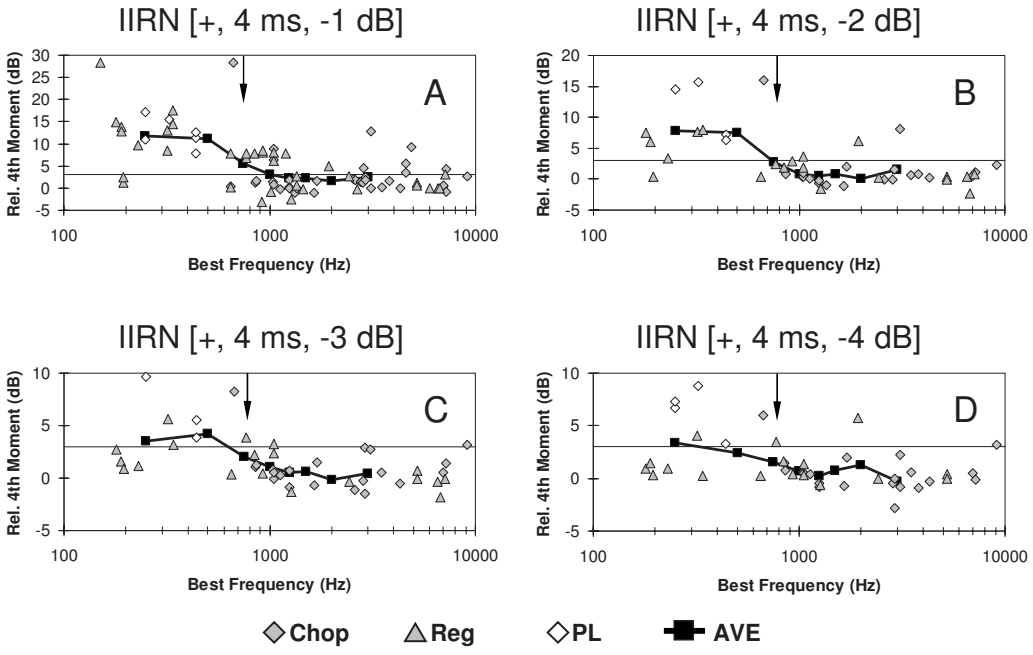


FIG. 7. Scatter plots illustrating the relative fourth moment as a function of BF for the Chopper group for delayed noise attenuations of −1 dB (A), −2 dB (B), −3 dB (C), and −4 dB (D). Gray diamonds show units with chopper PST histograms (Chop); gray triangles show units characterized as regular (Reg). The open diamonds illustrate the relative fourth moments obtained from low BF phase-locked (PL) units that could not be characterized as either primarylike or chopper. The black squares and solid black line show the relative fourth moment averaged in one-octave bands (ave), centered at harmonic frequencies of the rippled noise (i.e., 250, 500, 750, 1000 Hz, etc). The downward pointing arrow indicates the frequency of the third harmonic. Note that the ordinate changes scales in (A)–(D).

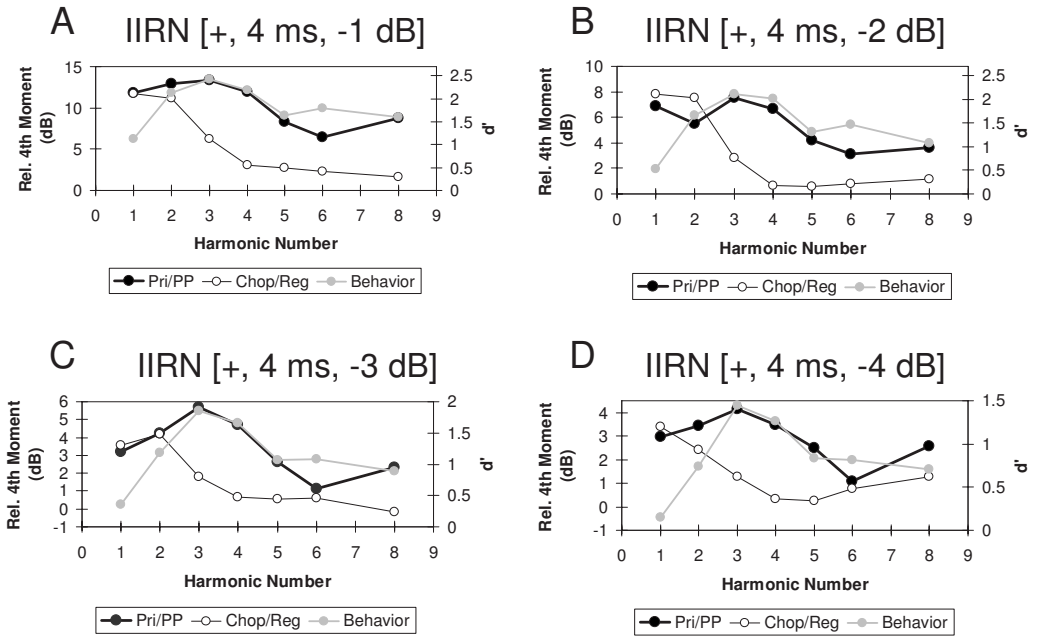


FIG. 8. Comparison of the relative fourth moment averaged in one-octave bands as a function of the harmonic number of the IIRN fundamental frequency (i.e., 250 Hz) for delayed noise attenuations of −1 dB (A), −2 dB (B), −3 dB (C), and −4 dB (D). These are the same functions illustrated in Figs. 6 and 7 by the solid black squares and lines. For comparison, behavioral data measured as d' from Shofner and Yost (1997) are shown by the gray circles and gray lines. The profiles of the averaged relative fourth moment for Primarylike units (black circles and black heavy lines) show more of a bandpass profile with a peak around the third harmonic and have a similar shape to the profile of the behavioral data. The profiles of the averaged relative fourth moment for Chopper units (open circles and thin black lines) have a lowpass profile and do not have a similar shape to the profile of the behavioral data. Note the change in the ordinate scale across the panels.

quently, the averages for each harmonic number cannot be treated as independent groups, thus making the use of a statistical analysis such as ANOVA across harmonics invalid. In order to compare neural profiles for the Primarylike and Chopper groups, the following correlational analysis was carried out.

Figure 8 also shows the behavioral performance for discrimination of bandpass filtered IIRN from bandpass filtered WBN averaged from five chinchillas (Shofner and Yost, 1997). Behavioral performance was measured as d', and the ordinate for the behavioral performance has been adjusted arbitrarily to align the behavioral data with the Primarylike data in each panel. In general, there appears to be a correlation between the shape of the averaged profile for the Primarylike group and the shape of the profile for the behavioral performance for all IIRNs illustrated. On the other hand, the neural profile for Chopper units and the profile for behavioral performance do not appear to be similar in shape at any delayed noise attenuation. These comparisons suggest that the shapes of the neural profiles for the Primarylike group are correlated with behavioral performance, whereas the profiles for the Chopper group are not correlated with behavioral performance. These relationships are quantified in Fig. 9, which shows scatter plots of d' as a function of the relative fourth moments averaged in the one-octave bands. For the Primarylike units, there is a significant correlation between d' and the averaged relative fourth moment (r=0.679; p =0.000036, one-tailed t-test). In contrast, there is not a significant correlation between d' and the relative fourth moment for the Chopper group of units (r=0.153; p=0.218, one-tailed t-test).

IV. DISCUSSION

A. General comments

A spectral dominance region has been described for pitch discrimination of rippled noises in human listeners (Bilsen and Ritsma, 1970; Yost and Hill, 1978; Yost, 1982;

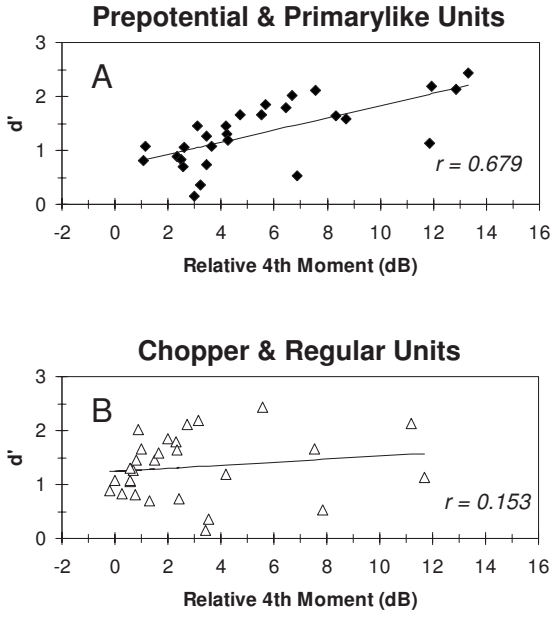


FIG. 9. Scatter plots showing the relationship between the relative fourth moment averaged in the one-octave bands, with d' obtained behaviorally for IIRN filtered by one-octave bandpass filters. Correlation is high for Primarylike units (A) but not for Chopper units (C). Correlation coefficients (r) are indicated in each panel.

Leek and Summers, 2001) as well as for the discrimination of IIRN from WBN in chinchillas (Shofner and Yost, 1997). These studies indicate that the region most effective in generating the pitch percept is generally around the third to fifth harmonic peaks of the rippled noise. The present results show that a neural correlate of the spectral dominance region exists in the temporal discharge of single units at the level of the cochlear nucleus for 4 ms delayed IIRNs.

The present study measured the temporal discharge patterns of cochlear nucleus units to IIRN and WBN using neural autocorrelograms. Presumably, a difference in the neuronal firing patterns evoked by IIRN and WBN underlies the behavioral discrimination in chinchillas (Shofner and Yost, 1997). Although IIRNs do not have periodic waveforms, they do possess temporal regularities that do not repeat in a periodic manner. Thus, rippled noises have been termed pseudoperiodic or quasiperiodic. The discharge patterns of Primarylike and Chopper units can reflect this temporal regularity as observed in the neural autocorrelograms. For both Primarylike and Chopper groups of units, there can be large differences in the temporal regularity for autocorrelograms obtained in response to IIRNs and WBN. That is, there is a regularity in the temporal discharge related to the delay that can be observed in response to the IIRNs. WBN does not possess any temporal regularity, and thus, there is no regularity in the temporal discharge of cochlear nucleus units. The difference in discharge patterns obtained in response to IIRNs and to WBN is reflected in the relative fourth moment, and presumably, this difference in discharge patterns is important in coloration discrimination (see below).

The results of the present study indicate that the spectral dominance region described for coloration discrimination in chinchillas (Shofner and Yost, 1997) is not found for units that are characterized by chopper and regular discharge patterns. The Chopper group of units show large differences in temporal discharge patterns between IIRN and WBN responses if the BFs correspond to the first and second harmonics of the IIRN. However, the shape of the profile of the relative fourth moment averaged in one-octave bands for these Chopper units is not correlated with the shape of the profile of behavioral performance. In contrast, the dominance region is better represented across BFs in the temporal discharge of cochlear nucleus units characterized by prepotentials in the spike waveforms and primarylike PST histograms. For this Primarylike group of units, the shape of the neural profile and the shape of behavioral performance are significantly correlated.

The approach of the present study was to examine the neural profile of the phase-locked activity across BFs for Primarylike and Chopper units in response to a fixed stimulus. The fixed stimuli were 4 ms delayed IIRN having four different delayed noise attenuations; these stimuli evoke a fixed pitch percept of 250 Hz that differ in pitch strength as delayed noise attenuation varies. Similar to the approach described for the representation of steady-state vowels in the cochlear nucleus (Blackburn and Sachs, 1990; Winter and Palmer, 1990b), data for the present study were pooled across animals to give an estimate of the population response. Unlike the vowel studies in which stimuli were presented at fixed sound levels, the present study presented stimuli at fixed levels above the threshold. In order to generate a meaningful estimate of the temporal discharge pattern through the autocorrelogram, it is important to evoke a high enough firing rate to collect a sufficient number of intervals over the 100 stimulus presentations. However, synchrony to the modulation frequency of sinusoidal-amplitude modulated tones decreases as sound level increases above the threshold for single-units in the cochlear nucleus (Rhode and Greenberg, 1994). Thus, at sound levels where the average firing rate is high, synchrony is reduced. A presentation level of 20 dB above the threshold was chosen to evoke a strong synchronous response and at the same time to ensure that a sufficient number of spike intervals for the autocorrelogram analysis were generated. Since the neural profile for Chopper units was not correlated with the behavioral profile of performance under these "optimal" stimulus conditions, it strongly suggests that dominance is not encoded in the temporal discharge of stellate cells.

Previous studies investigating physiological correlates of the dominance region have been carried out using complex tones. Greenberg *et al.* (1987) recorded the FFR from human listeners to two-tone complexes of successive harmonics of a fundamental frequency and showed that the largest responses were obtained for the third and fourth paired harmonics. The FFR is an auditory evoked potential based on the synchronous discharge of brainstem auditory neurons (Marsh *et al.*, 1972). Cariani and Delgutte (1996) measured ISIs for single auditory nerve fibers in response to harmonic tone complexes comprised of the 3rd–12th harmonics. In this tone complex, the third to fifth harmonics were of one fundamental frequency, whereas the 6th–12th harmonics were of another fundamental frequency. The predominant ISI corresponded to the fundamental frequency of harmonics 3–5 when the fundamental was 160 Hz or higher and thus showed a neural correlate of the dominance region. In response to iterated rippled noise with negative gain, Sayles and Winter (2007) found examples of both primarylike and chopper units in guinea pigs that showed synchronous responses related to the ambiguous pitches perceived by human listeners for units having BFs corresponding to $4/d$, where d ranged from 4 to 16 ms. These BFs corresponded to the dominance region. For the negative gain iterated rippled noises used in the above guinea pig study, the pitch that is evoked in human listeners would not be ambiguous but would be one octave lower than $1/d$, because a large number of iterations was used (i.e., 16 iterations) [see Yost (1996a)]. It is unknown whether the guinea pig perceives the pitch to be ambiguous or at the lower octave for these particular rippled noises. Moreover, for chopper units, the most prominent interspike interval did not necessarily correspond to the ambiguous pitch but rather could correspond to the natural chopping frequency of the unit [see Fig. 3 in Sayles and Winter (2007)]. Thus, these findings suggest that there may be differences in the representation of the dominance region between chopper units and primarylike units.

The results of the present study further extend these findings into the cochlear nucleus for rippled noise stimuli. The delay used in the present study was fixed at 4 ms, and

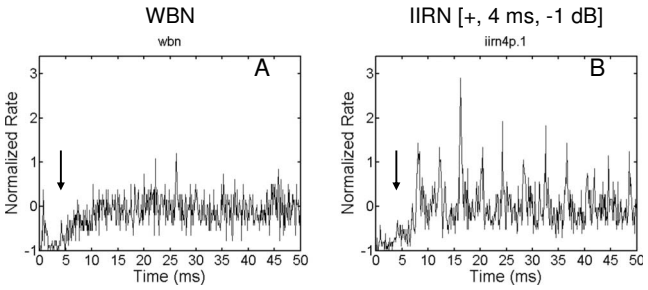


FIG. 10. Normalized autocorrelograms obtained from a unit in the Chopper group with a BF of 770 Hz in response to WBN (A) and IIRN [+, 4 ms, −1 dB) (B). Arrows indicate the location of the 4 ms ISI.

direct comparisons can be made between the observed physiological responses and behavioral performance obtained from the same species with the same stimuli (Shofner and Yost, 1997). The findings of the present and previous physiological studies (Greenberg *et al.*, 1987; Cariani and Delgutte, 1996; Sayles and Winter, 2007) argue that the neural representation of the dominance region of pitch is established early on in auditory processing and is not the result of higher level processing mechanisms. It is interesting to note that all of these physiological studies show the existence of the spectral dominance region based on the temporal discharge patterns of auditory neurons.

B. Use of the fourth moment

In the present study, the responses of cochlear nucleus units were analyzed in the context of the behavioral data previously obtained (Shofner and Yost, 1997). In that behavioral task, chinchillas discriminated an IIRN from WBN, and presumably, this discrimination is based on a difference between neuronal firing patterns in response to IIRN and those in response to WBN. It is the difference in discharge patterns obtained in response to 4 ms IIRNs and WBN that the present study attempts to quantify. One approach to this problem might be to generate period histograms and compute synchronization index or vector strength. However, as previously described, IIRNs are not truly periodic stimuli, but rather their waveforms have regularities that are not repeated in a periodic manner. Thus, since WBN and IIRN are not periodic stimuli, period histograms cannot be utilized.

Alternatively, differences in temporal discharge patterns could be assessed using ISI statistics. For example, the normalized rate at a time lag of 4 ms is related to the total number of 4 ms ISIs and could potentially be used to quantify the difference in discharge patterns between IIRN responses and WBN responses. However, any measure based exclusively on the number of 4 ms ISIs makes the assumption that in the behavioral discrimination, the decision maker knows beforehand to analyze 4 ms ISIs. This assumption may or may not be true. More importantly, however, units can show a difference in discharge patterns between 4 ms IIRN and WBN without showing a change in the number of 4 ms ISIs.

Figure 10 illustrates normalized autocorrelograms obtained in response to WBN and IIRN [+, 4 ms, −1 dB] for a 770 Hz BF unit classified as a Chopper. Note that the normalized rate at a time lag of 4 ms is small for both responses (arrows in Fig. 10, and there is no clear difference between the normalized rate at 4 ms for the WBN and IIRN responses. The normalized rate at 4 ms is −0.304 for the WBN response and −0.316 for the IIRN response. Thus, it is unlikely that this unit can encode the difference between 4 ms IIRN and WBN based only on the number of 4 ms ISIs. Nevertheless, there is a clear difference in the temporal discharge patterns between the responses to the IIRN and WBN. That is, this unit clearly responds differently to the IIRN than it does to the WBN even though there is no apparent difference in the number of 4 ms ISIs.

The fourth moment of a waveform is related to the fluctuations in instantaneous power (Hartmann, 1998), and it is clear in the example illustrated in Fig. 10 that the fluctuations in the response to IIRN are greater than those for the WBN. Strictly speaking, the fourth moment does not require the response to be synchronous, and large fourth moments could be obtained if the neural discharge pattern to IIRN showed deviations that were unrelated to the delay. However, the fact that the neural discharge patterns obtained in response to IIRNs often show periodicities related to the delay does not preclude the use of the fourth moment as an analytical tool. The use of the fourth moment was chosen because it captures the difference in the discharge patterns between IIRNs and WBN, and it makes no *a priori* assumptions that the decision maker must specifically analyze 4 ms ISIs.

C. Correlates of pitch strength

Iterated rippled noise evokes the perception of two simultaneous sound sources in human listeners: (1) the perception of a noise source and (2) the perception of a tonal source (Patterson *et al.*, 1996). The relative amounts of these two perceptions are determined by the number of iterations or the delayed noise attenuation. As the number of iterations increases (Yost, 1996b) or as the amount of delayed noise attenuation decreases (Shofner and Selas, 2002), there is a progressive decrease in the noise percept and a progressive increase in the tonal percept. That is, the saliency of the pitch percept (i.e., pitch strength) increases. Thus, the pitch strength of iterated rippled noise can be made to vary from that of a flat-spectrum noise to that of a pulse train. Chinchillas trained to discriminate IIRN [+, *d*, −1 dB] from WBN and tested in a stimulus generalization paradigm with IIRNs having the same delays but varying delayed noise attenuations show increases in behavioral responses as delayed noise attenuation decreases (Shofner *et al.*, 2005). These generalization gradients obtained from chinchillas presumably reflect a psychological dimension related to the delayed noise attenuation (i.e., pitch strength).

Physiological correlates of pitch strength have been described by Griffiths *et al.* (2001). As the number of iterations of the rippled noise increases, there is an increase in blood flow in each of the central auditory nuclei, including the cochlear nucleus, as measured by functional magnetic resonance imaging in the human auditory system. This finding suggests that there is an increase in neural activity as the number of iterations increases and is a correlate of the in-

crease pitch strength. Similarly, the results of the present study show that there is an increase in the periodicity in the discharge patterns of Primarylike and Chopper units as the delayed noise attenuation decreases (e.g., see Figs. 4 and 5). The ANOVA indicated significant effects of delayed noise attenuation for both Primarylike and Chopper groups for units with BFs less than or equal to 1250 Hz (i.e., less than or equal to the fifth harmonic of the IIRN). Thus, the increase in pitch strength observed with the decrease in delayed noise attenuation appears to be related to an increase in the strength of periodicity in the temporal discharge pattern of cochlear nucleus units.

D. Roles of primarylike and chopper units

Several studies have shown that the phase-locked responses to amplitude modulation are larger for chopper units than for primarylike units (Frisina *et al.*, 1990; Rhode and Greenberg, 1994; Wang and Sachs, 1994; Shofner *et al.*, 1996). Morphologically, chopper units correspond to stellate (multipolar) cells (Rhode *et al.*, 1983; Rouiller and Ryugo, 1984; Smith and Rhode, 1989; Ostapoff *et al.*, 1994), which send an axonal projection directly to the inferior colliculus (Adams, 1979; Cant, 1982). The enhanced ability of chopper units to encode amplitude modulation has suggested the importance of these units in processing information related to pitch. Indeed, neural models of periodicity processing have been based heavily on contributions from chopper units (Borst *et al.*, 2004; Meddis and O'Mard, 2006; Dicke *et al.*, 2007).

Although, both primarylike and chopper units with low BFs can encode temporally the difference in pitch of iterated rippled noises generated with positive and negative gains (Shofner, 1999; Verhey and Winter, 2006; Sayles and Winter, 2007), chopper units can also show bandpass tuning to iterated rippled noise delay as estimated from first-order ISI historgrams that is independent of the stimulus level (Wiegrebe and Winter, 2001). This latter finding has also been used to argue for the importance of chopper units in processing pitch-related information. The above discussion suggests that chopper units have an important role in periodicity processing.

The results of the present study show that a representation of the dominance region of pitch for 4 ms delayed IIRN does not appear to exist in the temporal discharge of the Chopper group of units. This finding should not be interpreted to imply that Chopper units are unimportant for periodicity processing. On the contrary, the present results show that these units give large averaged relative fourth moments if the BFs are around the first and second harmonics. For the 4 ms delayed IIRNs used in the present study, these harmonics would correspond to units having BFs below 500 Hz. In other words, low BF Chopper units show large differences in discharge patterns between 4 ms IIRN and WBN. Higher BF Chopper units, however, do not show as large of a difference in discharge patterns between IIRN and WBN, as evidenced by their smaller averaged relative fourth moments at harmonics above 2. Consequently, the dominance region is not well represented in these units.

The lack of a neural representation of the dominance region for the Chopper group presumably reflects the poor phase-locking ability of chopper units to BF tones. At 1000 Hz the synchronization index is approximately 0.2 for chopper units but is around 0.6 for primarylike units in the guinea pig (Winter and Palmer, 1990a). In a cat for 1000 Hz tones, the synchronization indices for chopper and primarylike units are approximately 0.4 and 0.8, respectively (Blackburn and Sachs, 1989). Thus, there is a decreased ability of chopper units to encode the frequency corresponding to the fourth harmonic of the 4 ms delayed IIRNs used in the present study. As a consequence, Chopper units with BFs corresponding to the third to fifth harmonics are unable to encode the 4 ms delay of the IIRNs as well as Primarylike units. Although chopper units can show enhanced phase-locking to envelope modulation, they show poorer phase-locking to the fine structure. IIRN is not highly modulated, but rather the delay information is predominately found in the fine structure [see Fig. 1 of Shofner and Selas (2002), for example]. Thus, the encoded neural information used by the animal for discriminating the 4 ms IIRNs from WBN is likely to be dominated by the discharge patterns of primarylike units rather than chopper units.

Although chopper units provide a synchronous neural discharge to the inferior colliculus, which is important for periodicity processing, the present findings of a neural correlate of the dominance region in the temporal discharge of primarylike units suggest that periodicity processing related to pitch perception is not exclusive to chopper units. In this regard, it is interesting to note that an alternative neural model based exclusively on the information encoded in primarylike units can also account for many aspects of periodicity processing (Nelson and Carney, 2004). Each of the neural models previously cited in this paper has attempted to account for aspects of periodicity processing through the exclusive use of either chopper units or primarylike units. As previously mentioned, chopper units correspond to stellate cells, which send a direct axonal projection to the inferior colliculus. Primarylike units correspond morphologically to bushy cells, and the information encoded in the discharge patterns of bushy cells will arrive at the inferior colliculus indirectly through projections from the superior olive [e.g., Oliver *et al.* (1995)]. Thus, the information encoded in the discharge patterns of primarylike and chopper units will ultimately converge at the level of the inferior colliculus.

In order to address how this convergence of neural information might affect the dominance region, the information for Primarylike and Chopper units having BFs within each of one-octave bands was evaluated using a signal detection theory approach. The d' for each one-octave band was computed as

$$d'_e = \frac{\mu_{\mathrm{Pr}\ i} - \mu_{\mathrm{Chop}}}{0.5(\sigma_{\mathrm{Pr}\ i} + \sigma_{\mathrm{Chop}})}, \tag{5}$$

where $\mu_{\mathrm{Pr}\ i}$ and μ_{Chop} are the averaged relative fourth moments in each one-octave band for the Primarylike and Chopper groups and $\sigma_{\mathrm{Pr}\ i}$ and σ_{Chop} are the standard deviations of the samples within each one-octave band. Note that

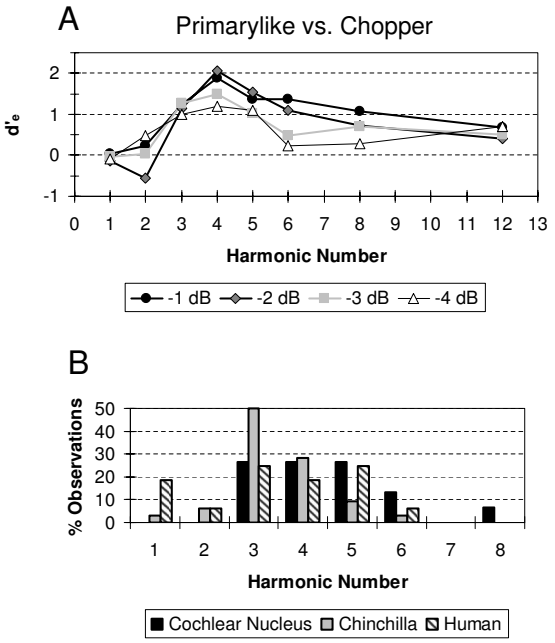


FIG. 11. (A) Profile of the dominance region expressed as d'_e as a function of harmonic number. d'_e represents the difference between Primarylike and Chopper groups (see text for details). (B) Black-filled bars show the distribution of harmonics, which resulted in a $d'_e \geq 1$ from the functions in (A). Gray-filled bars and striped bars show the distributions obtained from a similar analysis for chinchillas (data from Shofner and Yost, 1997) and normal hearing human listeners [data from Leek and Summers (2001)], respectively.

Kim *et al.* (1990) showed that Eq. (5) is equivalent to d'_e as defined by Green and Swets (1988) (p. 98). This analysis treats the information in the Primarylike group as the signal and the information in the Chopper group as the noise.

The d'_e in each one-octave band is illustrated in Fig. 11(A) where the octave bands are shown as harmonic numbers. Note that across the four delayed noise attenuations (i.e., −1, −2, −3, and −4 dB), the third to fifth harmonics consistently have d'_e values greater than or equal to 1. The harmonics that showed a d'_e greater than or equal to 1 were determined for each of the delayed noise attenuations [Fig. 11(B)]. For comparison, Fig. 11(B) shows similar data obtained from chinchillas discriminating 4 ms IIRN from WBN (Shofner and Yost, 1997) and normal hearing human listeners discriminating a 4 ms iterated rippled noise standard from another rippled noise having a 12% pitch change (Leek and Summers, 2001). This analysis suggests that the neural profile of the dominance region becomes well defined when the information in the discharge patterns of both bushy cells and stellate cells is combined.

In summary, the results of the present study argue for the importance of cochlear nucleus bushy cells in processing periodicity information related to pitch. It should be emphasized that this is not meant to imply that this is the exclusive pitch pathway. While it is likely that other cochlear nucleus cell types also contribute to the overall pitch perception, the contributions of bushy cells to processing pitch-related information should not be ignored.

ACKNOWLEDGMENTS

The research was supported by the National Institute on Deafness and Other Communication Disorders Grant R01 DC005596.

Abeles, M. (**1982**). "Quantification, smoothing, and confidence limits for single-units' histograms," J. Neurosci. Methods **5**, 317–325.
Adams, J. C. (**1979**). "Ascending projections to the inferior colliculus," J. Comp. Neurol. **183**, 519–538.
Amagai, S., Dooling, R. J., Shamma, S., Kidd, T. L., and Lohr, B. (**1999**). "Detection of modulation in spectral envelopes and linear-rippled noises by budgerigars (Melopsittacus undulatus)," J. Acoust. Soc. Am. **105**, 2029–2035.
Bassett, I. G., and Eastmond, E. J. (**1964**). "Echolocation: Measurement of pitch versus distance for sounds reflected from a flat surface," J. Acoust. Soc. Am. **36**, 911–916.
Bilsen, F. A., and Ritsma, R. J. (**1970**). "Some parameters influencing the perceptibility of pitch," J. Acoust. Soc. Am. **47**, 469–475.
Blackburn, C. C., and Sachs, M. B. (**1989**). "Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis," J. Neurophysiol. **62**, 1303–1389.
Blackburn, C. C., and Sachs, M. B. (**1990**). "The representations of the steady-state vowel sound /ɛ/ in the discharge patterns of cat anteroventral cochlear nucleus neurons," J. Neurophysiol. **63**, 1191–1212.
Borst, M., Langner, G., and Palm, G. (**2004**). "A biologically motivated neural network for phase extraction from complex sounds," Biol. Cybern. **90**, 98–104.
Cant, N. B. (**1982**). "Identification of cell types in the anteroventral cochlear nucleus tht project to the inferior colliculus," Neurosci. Lett. **32**, 241–246.
Cant, N. B., and Morest, D. K. (**1979**). "The bushy cells in the anteroventral cochlear nucleus of the cat. A study with the electron microscope," Neuroscience **4**, 1925–1945.
Cariani, P. A., and Delgutte, B. (**1996**). "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch," J. Neurophysiol. **76**, 1717–1734.
Cohen, M. A., Grossberg, S., and Wyse, L. L. (**1995**). "A spectral network model of pitch perception," J. Acoust. Soc. Am. **98**, 862–879.
Dicke, U., Ewert, S. D., Dau, T., and Kollmeier, B. (**2007**). "A neural circuit transforming temporal periodicity information into a rate-based representation in the mammalian auditory system," J. Acoust. Soc. Am. **121**, 310–326.
Frisina, R. D., Chamberlain, S. C., Brachman, M. L., and Smith, R. L. (**1982**). "Anatomy and physiology of the gerbil cochlear nucleus: An improved surgical approach for microelectrode studies," Hear. Res. **6**, 259–275.
Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (**1990**). "Encoding of amplitude modulation in the gerbill cochlear nucleus: I. A hierarchy of enhancement," Hear. Res. **44**, 99–122.
Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos, CA).
Greenberg, S., Marsh, J. T., Brown, W. S., and Smith, J. C. (**1987**). "Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones," Hear. Res. **25**, 91–114.
Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O., and Patterson, R. D. (**2001**). "Encoding of the temporal regularity of sound in the human brainstem," Nat. Neurosci. **4**, 633–637.
Guinan, J. J., and Li, R. Y.-S. (**1990**). "Signal processing in brainstem auditory neurons which receive giant endings (calyces of Held) in the medial nucleus of the trapezoid body of the cat," Hear. Res. **49**, 321–334.
Hartmann, W. M. (**1998**). *Signals, Sound, and Sensation* (Springer-Verlag, New York).
Houtgast, T. (**1977**). "Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker," J. Acoust. Soc. Am. **62**, 409–415.
Kim, D. O., Chang, S. O., and Siriani, J. G. (**1990**). "A population study of auditory-nerve fibers in unanesthetized decerebrate cats: Responses to pure tones," J. Acoust. Soc. Am. **87**, 1648–1655.
Leek, M. R., and Summers, V. (**2001**). "Pitch strength and pitch dominance of iterated rippled noises in hearing-impaired listeners," J. Acoust. Soc. Am. **109**, 2944–2954.
Lorente de No, R. (**1981**). *The Primary Acoustic Nuclei* (Raven, New York).
Marsh, J. T., Smith, J. C., and Worden, F. G. (**1972**). "Receptor and neural

responses to auditory masking of low frequency tones," Electroencephalogr. Clin. Neurophysiol. **32**, 63–74.
Meddis, R., and Hewitt, M. J. (**1991**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and O'Mard, L. P. (**2006**). "Virtual pitch in a computational physiological model," J. Acoust. Soc. Am. **120**, 3861–3869.
Morest, D. K., Hutson, K. A., and Kwok, S. (**1990**). "Cytoarchitectonic atlas of the cochlear nucleus of the chinchilla, Chinchilla laniger," J. Comp. Neurol. **300**, 230–248.
Nelson, P. C., and Carney, L. H. (**2004**). "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," J. Acoust. Soc. Am. **116**, 2173–2186.
Niemiec, A. J., Yost, W. A., and Shofner, W. P. (**1992**). "Behavioral measures of frequency selectivity in the chinchilla," J. Acoust. Soc. Am. **92**, 2636–2649.
Oliver, D. L., Beckius, G. E., and Shneiderman, A. (**1995**). "Axonal projections form the lateral and medial superior olive to the inferior colliculus of the cat: A study using electron microscopic autoradiography," J. Comp. Neurol. **360**, 17–32.
Osen, K. K. (**1969**). "Cytoarchitecture of the cochlear nuclei in the cat," J. Comp. Neurol. **136**, 453–484.
Ostapoff, E.-M., Feng, J. J., and Morest, D. K. (**1994**). "A physiological and structural study of neuron types in the cochlear nucleus. II. Neuron types and their structural correlation with response properties," J. Comp. Neurol. **346**, 19–42.
Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (**1996**). "The relative strength of the tone and noise components in iterated rippled noise," J. Acoust. Soc. Am. **100**, 3286–3294.
Pfeiffer, R. R. (**1966**). "Anteroventral cochlear nucleus: Wave forms of extracellularly recorded spike potentials," Science **154**, 667–668.
Pick, G. F. (**1980**). "Level dependence of psychophysical frequency resolution and auditory filter shape," J. Acoust. Soc. Am. **68**, 1085–1095.
Plack, C. J., Oxenham, A. J., Fay, R. R., and Popper, A. N. (**2005**). *Pitch: Neural Coding and Perception* (Springer-Verlag, New York).
Plomp, R. (**1967**). "Pitch of complex tones," J. Acoust. Soc. Am. **41**, 1526–1533.
Redd, E. E., Cahill, H. B., Pongstaporn, T., and Ryugo, D. K. (**2002**). "The effects of congenital deafness on auditory nerve synapses: Type I and type II multipolar cells in the anteroventral cochlear nucleus of cats," J. Assoc. Res. Otolaryngol. **3**, 418–429.
Rhode, W. S., and Greenberg, S. (**1994**). "Encoding amplitude modulation in the cochlear nucleus of the cat," J. Neurophysiol. **71**, 1797–1825.
Rhode, W. S., Oertel, D., and Smith, P. H. (**1983**). "Physiological response properties of cells labeled intracellularly with horseradish peroxidase in cat ventral cochlear nucleus," J. Comp. Neurol. **213**, 448–463.
Ritsma, R. J. (**1967**). "Frequencies dominant in the perception of the pitch of complex sounds," J. Acoust. Soc. Am. **42**, 191–198.
Rouiller, E. M., and Ryugo, D. K. (**1984**). "Intracellular marking of physiologically characterized cells in the ventral cochlear nucleus of the cat," J. Comp. Neurol. **225**, 167–186.
Ruggero, M. A., Rich, N. C., Robles, L., and Shivapuja, B. G. (**1990**) "Middle-ear response in the chinchilla and its relationship to mechanics at the base of the cochlea," J. Acoust. Soc. Am. **87**, 1612–1629.
Ryugo, D. F., and Fekete, D. M. (**1982**). "Morphology of primary axosomatic endings in the anteroventral cochlear nucleus of the cat: A study of the endbulbs of held," J. Comp. Neurol. **210**, 239–257.
Ryugo, D. K., and Sento, S. (**1991**). "Synaptic connections of the auditory nerve in cats: Relationships between end bulbs of Held and spherical bushy cells," J. Comp. Neurol. **305**, 35–48.
Sayles, M., and Winter, I. M. (**2007**). "The temporal representation of the delay of dynamic iterated rippled noise with positive and negative gain by single-units in the ventral cochlear nucleus," Brain Res. **1171**, 52–66.
Schreiner, C. E., and Calhoun, B. M. (**1994**). "Spectral envelope coding in the cat primary auditory cortex: Properties of ripple transfer functions," Aud. Neurosci. **1**, 39–61.
Sento, S., and Ryugo, D. K. (**1989**). "Endbulbs of held and spherical bushy cells in cats: Morphological correlates with physiological properties," J. Comp. Neurol. **280**, 553–562.
Shamma, S. A., Versnel, H., and Kowalski, N. (**1995**). "Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra," Aud. Neurosci. **1**, 233–254.
Shofner, W. P. (**1991**). "Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla," J. Acoust. Soc. Am. **90**, 2450–2466.
Shofner, W. P. (**1999**). "Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: Analysis of neural autocorrelograms," J. Neurophysiol. **81**, 2662–2674.
Shofner, W. P., and Selas, G. (**2002**). "Pitch strength and Stevens's power law," Percept. Psychophys. **64**, 437–450.
Shofner, W. P., Sheft, S., and Guzman, S. J. (**1996**). "Responses of ventral cochlear nucleus units in the chinchilla to amplitude modulation by low-frequency, two-tone complexes," J. Acoust. Soc. Am. **99**, 3592–3605.
Shofner, W. P., Whitmer, W. M., and Yost, W. A. (**2005**). "Listening experience with iterated rippled noise alters the perception of 'pitch' strength of complex sounds in the chinchilla," J. Acoust. Soc. Am. **118**, 3187–3197; **99**, 3592–3605.
Shofner, W. P., and Yost, W. A. (**1997**). "Discrimination of rippled-spectrum noise from flat-spectrum noise by chinchillas: Evidence for a spectral dominance region," Hear. Res. **110**, 15–24.
Smith, P. H., Joris, P. X., Carney, L. H., and Yin, T. C. T. (**1991**). "Projections of physiologically characterized globular bushy cell axons from the cochlear nucleus of the cat." J. Comp. Neurol. **304**, 387–407.
Smith, P. H., Joris, P. X., and Yin, T. C. T. (**1993**). "Projections of physiologically characterized spherical bushy cell axons from the cochlea nucleus of the cat: Evidence for delay lines to the medial superior olive," J. Comp. Neurol. **331**, 245–260.
Smith, P. H., and Rhode, W. S. (**1987**). "Characterization of HRP-labeled globular bushy cells in the cat anteroventral cochlear nucleus," J. Comp. Neurol. **266**, 360–375.
Smith, P. H., and Rhode, W. S. (**1989**). "Structural and functional properties distinguish two types of multipolar cells in the ventral cochlear nucleus," J. Comp. Neurol. **282**, 595–616.
Verhey, J. L., and Winter, I. M. (**2006**). "The temporal representation of the delay of iterated rippled noise with positive or negative gain by chopper units in the cochlear nucleus," Hear. Res. **216–217**, 43–51.
Wang, X., and Sachs, M. B. (**1994**). "Neural encoding of single-formant stimuli in the cat. II. Responses of anteroventral cochlear nucleus units," J. Neurophysiol. **71**, 59–78.
Wiegrebe, L., and Winter, I. M. (**2001**). "Temporal representation of iterated rippled noise as a function of delay and sound level in the ventral cochlear nucleus," J. Neurophysiol. **85**, 1206–1219.
Winter, I. M., and Palmer, A. R. (**1990a**). "Responses of single units in the anteroventral cochlear nucleus of the guinea pig," Hear. Res. **44**, 161–178.
Winter, I. M., and Palmer, A. R. (**1990b**). "Temporal responses of primary-like anteroventral cochlear nucleus units to the steady-state vowel /i/," J. Acoust. Soc. Am. **88**, 1437–1441.
Winter, I. M., Wiegrebe, L., and Patterson, R. D. (**2001**). "The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig," J. Physiol. (London) **537**, 556–566.
Yost, W. A. (**1982**). "The dominance region and ripple noise pitch: A test of the peripheral weighting model," J. Acoust. Soc. Am. **72**, 416–425.
Yost, W. A. (**1996a**). "Pitch of iterated rippled noise," J. Acoust. Soc. Am. **100**, 511–518.
Yost, W. A. (**1996b**). "Pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **100**, 3329–3335.
Yost, W. A., and Hill, R. (**1978**). "Strength of pitches associated with ripple noise," J. Acoust. Soc. Am. **64**, 485–492.
Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). "A time domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066–1078.
Young, E. D., Robert, J.-M., and Shofner, W. P. (**1988**). "Regularity and latency of units in ventral cochlear nucleus: Implications for unit classification and generation of response properties," J. Neurophysiol. **60**, 1–29.

Objective threshold estimation and measurement of the residual background noise in auditory evoked potentials of goldfish

Jianqiang Xiao[a)]
Department of Psychology, Hunter College, The City University of New York (CUNY), 695 Park Avenue, New York City, New York, 10021

Christopher B. Braun[b)]
Department of Psychology, Hunter College, The City University of New York (CUNY), 695 Park Avenue, New York City, New York 10021 and Biopsychology Program, The Graduate School and University Center (CUNY), 365 5th Avenue, New York City, New York 10016

(Received 19 October 2007; revised 31 July 2008; accepted 13 August 2008)

A survey of papers using auditory evoked potentials (AEPs) published over the last 10 years (Table I) demonstrates that most AEP studies in animals have used subjective methods for auditory threshold determination. Subjective methods greatly reduce the value of statistical hypothesis testing and jeopardize tests of hypothetical experimental group differences in hearing sensitivity. Correspondingly, many attempts have been made to develop objective threshold determination methods, but these have not been used widely. Further, they seldom include an appreciation of the effects of residual noise in the AEP. In this study, AEPs evoked by tonal and noise stimuli in goldfish (*Carassius auratus*) were recorded and the residual background noise was measured and analyzed in detail. High variability was found in residual noise, but can be effectively controlled with a simple modification of averaging routines. Considerable interobserver disagreements were found using subjective threshold estimation. An objective method of threshold determination was developed based on comparison between AEP amplitude and controlled residual noise, using a signal detection theory approach to set specific threshold criteria. The usefulness of AEP in hypothesis testing for auditory function requires more control over residual background noise amplitudes and the use of objective threshold determination techniques.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2982366]

PACS number(s): 43.64.−q, 43.64.Ri, 43.80.−n, 43.80.Lb [WPS] Pages: 3053–3063

I. INTRODUCTION

The surface recording of auditory evoked potentials (AEPs) utilizes synchronous averaging of far-field neuronal potentials evoked by repetitive presentations of auditory stimuli. The technique remains popular for many reasons, but especially because it is a noninvasive and time-efficient means of measuring auditory system function. It has now become widely used in clinical and basic hearing research in both human and diverse vertebrate animal models [mammals (e.g., Boettcher, 2002), birds (e.g., Brittan-Powell *et al.*, 2002), reptiles (e.g., Higgs *et al.*, 2002), amphibians (e.g., Katbamna *et al.*, 2006), and fishes (e.g., Kenyon *et al.*, 1998)].

Most recently, AEP audiometry has been used to statistically test a wide variety of hypotheses on the function, development, and evolution of audition. Examples include the effect of chemical and noise exposure on auditory sensitivity (e.g., Lu and Tomchik, 2002; Le Prell *et al.*, 2004; Popper *et al.*, 2005; Porter *et al.*, 2006; Cordova and Braun, 2007), ontogenetic changes in hearing abilities (e.g., Boettcher, 2002; Higgs *et al.*, 2003; Song *et al.*, 2006), or comparative studies of closely related species (e.g., Ramcharitar and Popper, 2004; Ramcharitar *et al.*, 2006). The vast majority of publications reporting AEP threshold estimates (in animals) in this and three other prominent journals over the last 10 years used subjective response determinations to estimate thresholds (Table I). In some publications, no mention of threshold determination technique at all was given. Typically, subjective response determination relies on visual inspection of the repeatability of replicate AEP averages using a subjective yes/no decision, occasionally based on expected minimal response amplitudes (e.g., Boettcher, 2002) and/or within a particular latency window (e.g., Ramcharitar and Popper, 2004). The heavy reliance on visual inspection techniques in the animal AEP literature is especially unfortunate because unsatisfactory interobserver agreements of visual inspection have been shown in several independent studies of experienced observers or clinicians (Arnold, 1985; Gans *et al.*, 1992; Vidler and Parker, 2004; Lv *et al.*, 2007). Unfortunately, the use of such unreliable subjective threshold determination techniques raises the possibility that hypothesis testing could be compromised by confirmation bias, particularly in the absence of double-blind protocols (Nickerson, 1998).

Because of the obvious shortcomings of the subjective threshold determination methods, attempts to objectively determine AEP threshold have been made repeatedly, espe-

[a)]Present address: Psychology Department, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854.
[b)]Author to whom correspondence should be addressed. Electronic mail: cbraun@hunter.cuny.edu

TABLE I. Summary of literature review. The table gives the number of studies using AEP for animal hearing threshold determination with different methods, including subjective methods of visual inspection and objective methods based on the measure of the response amplitude or correlation analysis. Some publications did not describe threshold determination techniques at all (not given). The nonexhaustive survey covered four journals for the years 1997–2006. HR: Hearing Research; JASA: The Journal of the Acoustical Society of America; JEB: The Journal of Experimental Biology; JCPA: Journal of Comparative Physiology A: (Neuroethology) Sensory, Neural, and Behavioral Physiology.

	HR	JASA	JCPA	JEB
Visual inspection	178	20	8	7
Response level	12	11	2	1
Correlation	1[a]		5[a]	
Not given	29			
Total	220	31	15	8

[a]One paper in each category used both correlation analysis and visual inspection; they are both counted only in the correlation category.

cially in clinical research (see Hyde *et al.*, 1998, for a review). Two general approaches to objective AEP threshold determination have been proposed: (1) Response repeatability measurement, i.e., a graphical correspondence should exist in responses evoked by repeat stimulus presentations, which can be measured with mathematical algorithms rather than judged by subjective visual inspection. The implementation of this method has been generally based on correlation analysis, either between AEP replicates evoked by the identical stimuli at the same level, or between recorded AEP and a predetermined response template (e.g., Elberling, 1979; Ozdamar *et al.*, 1994; Yan, 1998; Lu and Tomchik, 2002). (2) AEP amplitude analysis, wherein the amplitude (from either time or frequency domain) of the AEP is used to determine if a stimulus-evoked response is present. Studies employing response-amplitude-based threshold estimations mainly involve a combination of three steps: (i) Measurement of the response amplitude, usually in terms of peak-to-peak (e.g., May *et al.*, 2002) or root mean square (rms) (e.g., Elberling and Don, 1984) voltage in the time domain or power in the frequency domain (e.g., Supin *et al.*, 2001; Popper *et al.*, 2005); (ii) Estimation of the background noise remaining in the average (see below); and (iii) Definition of a criterion minimum AEP amplitude for an evoked response, such as 3 dB above the background noise (Popper *et al.*, 2005), two standard deviations (SDs) above the average background amplitude (May *et al.*, 2002), or simply using an arbitrary value with no stated relation to the background noise (e.g., 0.1–0.2 μV in Mitchell *et al.*, 1996). A simpler AEP amplitude threshold determination approach uses the relationship between response amplitude and stimulus level to determine which stimulus levels are linearly related to AEP amplitude, regardless of background noise amplitudes. The threshold level can then be defined as the zero-crossing value or inflection point, below which the response amplitude does not change monotonically with the stimulus level (e.g., Supin *et al.*, 2001; Brittan-Powell *et al.*, 2002).

AEP amplitude analysis, while providing much needed objectivity, cannot be simply applied if the experimental conditions change. The amplitude of any AEP average is a reflection of both the response and the residual background noise in an average (RBN); the latter reflects multiple sources, including other neural activity, muscular contractions, and ambient electrical pollution. When the stimulus level is just above threshold levels, the AEP amplitude is expected to be just greater than the RBN amplitude, thus when the background noise amplitude varies greatly, either spontaneously or as a result of experimental condition [such as with or without administration of neuromuscular blocker (Kenyon *et al.*, 1998)], the estimate of threshold based on response amplitude will be contaminated by variability in noise amplitude, accordingly. Thus, AEP amplitudes cannot be directly compared without concurrent measures of RBN amplitude. Any estimates of response strength will be contaminated unless RBN amplitude differences are also considered.

Residual noise amplitudes in an AEP average roughly reflect the number of signal sweeps composing the average AEP, with a theoretical relationship of $BN/\sqrt{N}$, where BN is the instantaneous background noise amplitude and N is the number of sweeps (Wong and Bickford, 1980). This relationship holds only if the background noise is stationary, so collecting a fixed number of sweeps [e.g., 500 sweeps by Katbamna *et al.* (2006), 1000 sweeps by Brittan-Powell *et al.* (2002), and 2000 sweeps by Kenyon *et al.* (1998)] does not ensure that RBN will be the same in each average. The instantaneous noise amplitude may vary across trials depending on the status of the test subject and the experimental setup. More damagingly, episodic noise from different sources can change the noise amplitude during averaging, weakening the relationship between sweep number and RBN amplitude. Therefore, objective response-amplitude threshold determination techniques should be combined with the control of RBN amplitude, especially in studies using statistical hypothesis testing. Bayesian averaging techniques may also be used to differentially weight individual sweeps (or blocks of sweeps) in an average (Eberling and Wahlgreen, 1985).

The RBN amplitude can be measured in many ways. Some authors have proposed to directly measure the amplitude (e.g., rms amplitude) of averages composed of an equal number of no-stimulus sweeps (e.g., Mann *et al.*, 2001) or of averages within separate prestimulus or postresponse windows of the stimulus sweep (e.g., May *et al.*, 2002). This approach may be a good approximation of the RBN amplitude presented in AEP averages containing responses, however, it does require additional data collection, and it may also be inaccurate if ongoing noise amplitude changes greatly during the course of data collection.

To circumvent these complications, other authors have proposed methods of directly estimating the RBN amplitude in AEP averages that also contain a response. Elberling and Don (1984) recorded the instantaneous voltage of a single time-point from each sweep composing the AEP average. If the noise is random and normally distributed, the standard deviation of these instantaneous voltages should be equal to the rms amplitude of the RBN embedded in the AEP average, after being corrected by the total number of sweeps. The ratio of the variance of the AEP to the single-point variance

can be used as an F statistic, which allows for the estimation of thresholds within specific confidence windows using approximated degrees of freedom (Elberling and Don, 1984). In their subsequent studies, Don and Elberling and colleagues showed that this method is also an effective means of estimating noise amplitudes in human audiometric data and can be used to determine how many sweeps should be included in an average (Don *et al.*, 1984; Don and Elberling, 1994; 1996). In the present report, the estimated RBN amplitude using this method will be presented as the RBN_{SP}.

Signal averaging in AEP measurements is based on the assumptions that the neural potential evoked by the repetitive stimulus is identical in each sweep, and the averaging process only reduces the amplitude of noise, i.e., potentials that do not have a fixed time relationship with the stimulus presentation. Based on this idea, an alternate method of RBN amplitude estimation was proposed by Wong and Bickford (1980), following Schimmel (1967). The RBN in an average AEP was estimated from the plus-minus (±) average, the average of all sweeps following a polarity-reversal of one half of the recorded sweeps. This ± average represents the same sweeps composing the AEP average, except that all time-locked features have been effectively removed by the polarity reversal of half of the sweeps. The rms of this ± average is an indirect estimate of the RBN amplitude in the AEP average (assessed from the same sweeps) and is termed as the $RBN_{\pm}$ hereafter.

The present report describes the measurement of RBN in goldfish AEP averages, and provides the range and variability of RBN amplitudes in multiple replicates of AEP tests. We also report on the interobserver consistency of threshold determination using visual inspection. Subsequently, we present an evaluation of RBN_{SP} and $RBN_{\pm}$ and conclude that $RBN_{\pm}$ is an effective and simple means of noise estimation. Finally, we describe an objective method of threshold determination based on comparing the AEP amplitude and RBN_{SP} with fixed RBN amplitudes. We also show how the principles of signal detection theory can be used to estimate the equivalent hit and false alarm rates of specific AEP threshold criteria.

II. GENERAL METHODS

Goldfish (*Carassius auratus*, comet variety) [N=21; standard length: 10±0.8 (SD) cm; body mass: 39±11.4 g] were obtained from a commercial distributor, maintained in filtered aquaria under 12L:12D light cycle and fed commercial food. The protocol of animal use in the study was approved by the Institutional Animal Care and Use Committee of Hunter College, CUNY.

Each animal was lightly anesthetized in 100 mg L^{-1} MS222 (3-aminobenzoic acid ethyl ester, Sigma) before an intramuscular injection with Flaxedil (gallamine triethiodide, Sigma; 5 μg g^{-1} body weight). The animal was then held in a sling constructed of a foam-lined aluminum strip strung from the sides of the tank [6 mm acrylic: 45×40×15 (depth) cm]. The tank contained 10 cm deep water and the animal was suspended with the top of skull above the water surface (no more than 2 mm). The exposed part was covered with a small piece of tissue paper to retain moisture. The animal was respirated by a gravity-fed water circulation system through a tube inserted in the mouth. To reduce the disturbance from environmental vibration and noise, the tank rested on a vibration-isolation table (Nano-K™, Minus K Technology, Inc., Inglewood, CA) within a 1.2×1.2×2 m single wall sound-attenuating booth (Controlled Acoustical Environments, Industrial Acoustics Company, Inc., New York, NY).

All stimuli presentation, data acquisition, and data management were operated via a TDT System 3 system (Tucker-Davis Technologies, Inc., Gainesville, FL; sampling frequency=24.4 kHz) controlled by customized routines in MATLAB® 7 and TDT RPvdsEx© 5.4.

Two stimulus types were presented to evoke the AEPs: 1000 Hz tone pips of 9 ms duration, with 2 ms linear rise/fall ramps, and band-limited (600–1400 Hz) white noise bursts of 7 ms duration with 0.01 ms linear ramps. These were presented with an interstimulus interval of 33.17 ms. The stimuli were generated by a TDT RP2.1 processor. Analog outputs were attenuated (TDT PA5), amplified (Crown D-75A), and broadcasted from a JBL Control® 29AV Speaker (frequency response: 45 Hz–18 kHz±3 dB), suspended ~60 cm above the water surface. To monitor the stimulus level, a B&K 8103 hydrophone was placed underwater adjacent to the fish. The hydrophone signal was fed through a B&K Nexus conditioning amplifier (with a 30–10 000 Hz band-pass filter) and digitized by the TDT RP2.1 processor. Estimation of tone pip level was based on averaged hydrophone signals, expressed as the largest rms value obtained over any 1-ms window (one stimulus cycle) during the burst [following Burkard (1984)]. To estimate noise-burst levels, ~2% of the individual stimulus sweeps were evenly sampled. The rms amplitude of each sweep was estimated over a 7-ms window, starting at stimulus onset. The stimulus level was calculated as the mean rms of all the samples and was expressed in dB re 1 μPa.

The AEPs were measured with a pair of silver-wire electrodes. The active electrode was positioned on the midline of the skull above the medulla region, and the reference one between the nostrils. The tank water near the fish was grounded to the ground of the headstage (TDT RA4LI). Electrical activity was amplified (TDT RA4PA preamplifier), then band-pass filtered (30–3000 Hz) and digitized with a TDT RA16 processor. Sweeps with peak-to-peak amplitude higher than 15 μV were rejected by an artifact rejection circuit. The average AEP responses were recorded over an 800-point (i.e., 32.8 ms long) window, beginning with stimulus onset.

III. EXPERIMENT 1: THRESHOLD ESTIMATION WITH UNCONTROLLED VARIATION IN RBN AMPLITUDE

A. Introduction

Repeated AEP tests were carried out to examine the variation in RBN amplitudes and the interobserver reliability of visual inspection for hearing threshold estimation.

B. Methods

Sixteen animals were tested with tone (N=10) and/or noise (N=15) bursts in this experiment. Stimulus level was decreased initially with steps of 10 dB (starting at 120–130 dB re 1 μPa) and then by 5 dB near threshold, until several steps below a provisionally estimated threshold. At each level, the stimuli were presented in four blocks of 500 sweeps, with alternating stimulus polarity by block (starting phases of 0° and 180°), forming a total of 2000 sweeps. These blocks were used to create two subaverages of 1000 sweeps each composed of an equal mix of both stimulus polarities. We defined threshold as the lowest stimulus level that evokes a response with all higher stimulus levels also evoking responses. The presence or absence of a response was judged by both visual inspection and by comparison of AEP and RBN amplitudes.

1. Visual inspection

Visual inspection of two replicate 1000 sweep subaverages for each single trial was performed by three experienced observers independently. The waveforms of the two subaverages were graphically overlapped and visually examined for their repeatability by each observer. The recorded signal was determined to contain a response when the two traces were judged to be repeatable.

2. Measurement of AEP and RBN amplitudes

The AEP rms amplitude was calculated over the first 16.4 ms (400 samples) rather than the whole sampling window (800 samples) from the average AEP of all 2000 sweeps. We adopted a shorter window because the duration of the response decreases as the stimulus level declines. Close to threshold, the response itself may only last a few milliseconds, so longer windows will underestimate the amplitude of near threshold AEP. We chose a 16.4-ms window as a compromise between accurate $RBN_{\pm}$ measurements (see below) and sensitive AEP amplitude measures. Noise amplitudes in the AEP were estimated as RBN_{SP} and $RBN_{\pm}$. To calculate RBN_{SP}, the instantaneous voltage at one fixed time-point (sample number 144, ~6 ms) was recorded in every sweep. The amplitude of the RBN presented in the average AEP was estimated as $RBN_{SP}=SD_{SP}/\sqrt{N}$, where SD_{SP} is the standard deviation of the instantaneous voltages and N is the total number of sweeps in the average (Elberling and Don, 1984). For tests with tonal stimuli (where stimulus polarity was alternated to reduce stimulus artifacts), the variances in instantaneous voltages were calculated individually from the sweeps of each stimulus polarity and averaged before calculating the overall SD_{SP}. To calculate $RBN_{\pm}$, we used the two 1000 sweep subaverages for each trial as described for visual inspection, each representing both stimulus polarities. One subaverage was subsequently reversed (multiplied by −1) before averaging with the other subaverage to generate the ± average. The $RBN_{\pm}$ was calculated as the rms of this ± average based upon all 2000 sweeps (Wong and Bickford, 1980). The AEP amplitude was then compared with the RBN amplitude estimate to determine the threshold (see below).

C. Results

Figure 1 shows pairs of representative mean AEP waveforms (each from 1000 sweeps) and the corresponding ± average noise waveforms (and the rms amplitude of each). Characteristic features can be seen from the AEP waveforms, such as declining amplitude, increasing latency and decreasing response duration with decreasing stimulus level, as have been observed before in this (Cordova and Braun, 2007) and other laboratories (e.g., Kenyon *et al.*, 1998; Lu and Tomchik, 2002).

1. Interobserver reliability of visual inspection

Each observer judged 25 detection thresholds using visual inspection of pairs of 1000 sweep averages at each stimulus level. The interobserver disagreement was profound. Cohen's Kappa values, which measure interobserver reliability (Cohen, 1960), were 0.76, 0.80, and 0.87 for the three pairs of observers. Complete agreement (all three observers in agreement) only occurred 16% of the time, and the three observers made entirely different decisions for 20% of the judgments. The average maximum difference between any two observers was 8 ± 6.1 dB for all the tests, and the average difference between all three observers was 6 ± 4.0 dB. Across all the tests, 28% of the mean differences between all three observers were at least two stimulus level-steps (≥10 dB), and only 52% of the average differences were within one level-step (5 dB).

When the observers made different judgments, the median judgment was reported as threshold for that stimulus level in that subject. Across animals, the mean detection threshold was 91 ± 3.2 dB re 1 μPa (N=10) for tonal stimuli and 92 ± 4.1 dB re 1 μPa (N=15) for noise stimuli.

2. Noise estimation accuracy

Since one of the basic assumptions behind the use of RBN_{SP} is that noise is normally distributed, we examined the distribution of instantaneous voltages to assess the validity of this assumption. In our recordings, instantaneous voltages frequently deviated from a normal distribution (judged by Kolmogorov–Smirnov tests). The single-point recordings in 221 of 1028 (21.5%) 500-value blocks were not drawn from normal distributions. The non-normal percentages were 21.7% and 21.4% for responses evoked by tonal and noise stimuli, respectively. Examining the dataset by individual subject, the portion of data that appears to contain non-normal noise varied from 9.6% to 50.0%, with a mean of $22.6\% \pm 13.18\%$. Further examination of the distribution of the single-point recordings showed that most of the non-normal 500-value blocks were not significantly skewed (84.6%) but were kurtotic. Of those blocks with significant kurtoses beyond a 95% confidence interval, the vast majority were leptokurtic (97.9%), rather than platykurtic (2.1%).

To further evaluate the accuracy of both noise estimates, we compared them to the rms amplitude of the AEP evoked by the stimuli at the lowest level presented in each series for all animals (at least 10 dB below threshold, mean difference from threshold was ~15 dB for tonal stimuli and ~28 dB for noise bursts). At these stimulus levels, the AEP may be

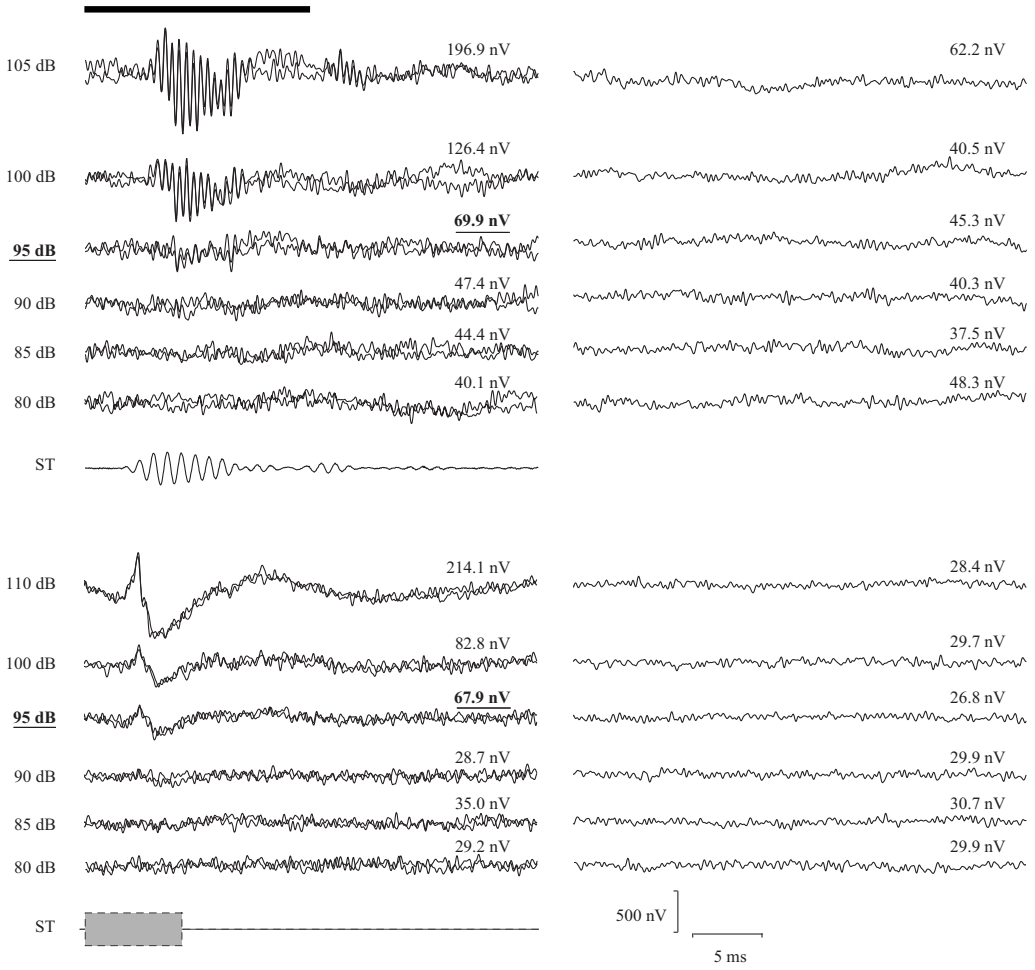


FIG. 1. Representative waveforms of average AEPs (two replicate averages of 1000 sweeps each are superimposed) evoked by tonal pips (top) and noise bursts (bottom). Corresponding ± average waveforms are plotted beside the average AEPs. Amplitude (rms) amplitudes of the average of all 2000 sweeps and of the ± average are given on the top of corresponding traces. Values on the left side of AEP traces are the sound pressure level of the evoking stimuli (re 1 μPa) and the threshold values are in bold font. A sample waveform of the tonal stimulus is given below the AEP waveforms. The schematic envelope of noise burst is also given as a rectangular shape. The time-window used to measure AEP amplitude is indicated as a heavy black bar at the top of the figure.

expected to contain only noise, and the rms of the average should be equal to the true RBN amplitude using the same sweeps. Therefore, we compared these no-response AEP rms values with the corresponding estimates obtained with both single-point and ± average methods. Since the evoked activity measured from the subject was filtered between 30 and 3000 Hz, but a 16.4-ms analysis window limits the noise to a bandwidth above 61 Hz (1/.0164), we recalculated the no-response AEP rms with the full 800-point window when comparing the AEP rms and the RBN_{SP}. Results of the comparisons showed no significant difference between the AEP rms and the $RBN_{\pm}$ (Wilcoxon matched-pairs signed-ranks test: $P=0.9158$, $N=25$). However, the RBN_{SP} estimates were significantly higher than the no-response AEP rms ($P<0.0001$, $N=25$). Similar results were found for both stimulus types: comparing $RBN_{\pm}$ and no-response AEP rms, $P=0.2324$ for tonal stimuli ($N=10$) and 0.3028 for noise bursts ($N=15$); using RBN_{SP} estimates, $P=0.0195$ (tone, $N=10$) and 0.0001 (noise, $N=15$). Based on these results, further analysis was based on $RBN_{\pm}$ rather than RBN_{SP}

3. Background noise variation

Across the 16 individual animals tested, the mean $RBN_{\pm}$ rms ranged between 32.2 and 54.6 nV after averaging 2000 sweeps. The overall average was 42.5 ± 6.96 nV rms ($N=16$), with a CV (coefficient of variation) of 16.4%. Obvious fluctuation in RBN amplitudes was observed both within individual test series and among series collected on different days (Fig. 2). The CV of RBN for each individual test day varied from 11.3% to 51.6% around a mean of $30.9 \pm 12.65\%$ ($N=16$).

4. Quantitative threshold estimation

The direct measure of RBN amplitude from AEPs recorded at very low stimulus levels (no response) was further used to estimate the error in $RBN_{\pm}$ measures. We calculated the mean ratio of the no-response AEP to $RBN_{\pm}$ amplitudes and its variance. If $RBN_{\pm}$ was identical to the true noise amplitude (as estimated by the no response condition), this ratio should equal one and its deviation from that value represents the degree of error in the $RBN_{\pm}$ estimate. In our data, this noise estimate error ratio (NER) did not significantly deviate from a normal distribution (Kolmogorov–Smirnov tests, $P=0.1855$, $N=16$). The distribution of NER (mean $=1.0$ and $SD=0.16$) can be used to estimate the distribution of AEP amplitudes at specific signal-to-noise ratios (SNRs). For example, with a 3 dB SNR in the AEP, 82.7% of average

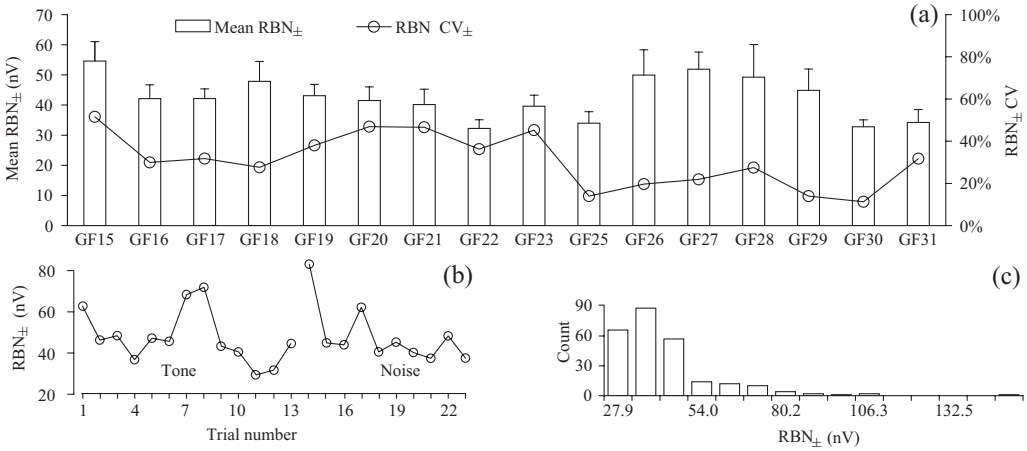


FIG. 2. Variation of RBN amplitude in experiment 1, illustrated with (a) interindividual variation of the mean $RBN_{\pm}$ (with error in SD) and CV of the $RBN_{\pm}$ after 2000 sweeps, (b) within-individual variation of $RBN_{\pm}$ obtained in consecutive trials from one representative animal, and (c) the frequency distribution of all $RBN_{\pm}$ levels in 2000 sweep averages collected from all 16 animals.

AEP amplitudes would be expected to be at least 20% larger than the corresponding RBN. Conversely, when there is no response present (AEP amplitude is equal to RBN on average: 0 dB SNR), only 10.6% of the individual AEP averages will be at least 20% greater than their corresponding RBN estimates. Thus, using a criterion of 120% of $RBN_{\pm}$ to define a threshold response, we correctly categorize over 80% of responses positively when the SNR in the AEP is at least 3 dB, with a false positive rate of approximately 10%.

Greater or lesser stringency may be achieved by using higher or lower criteria, respectively. Differences in NER SD between datasets will result in altered hit and false alarm rates, but a SD as high as 0.25 (while mean NER=1.0) still provides high detectability of 3 dB AEP signals (Fig. 3). Similarly, changes in the mean NER can also affect the hit and false alarm rates slightly, by increasing both rates with increasing mean NER and *vice versa* (figure not shown). However, dividing the present dataset into two subgroups of eight animals resulted in similar NER estimates (1.0 ± 0.15 and 1.0 ± 0.22), and other samples from unpublished datasets and from the same subjects (not shown) also produced very similar values (also see Sec. IV C).

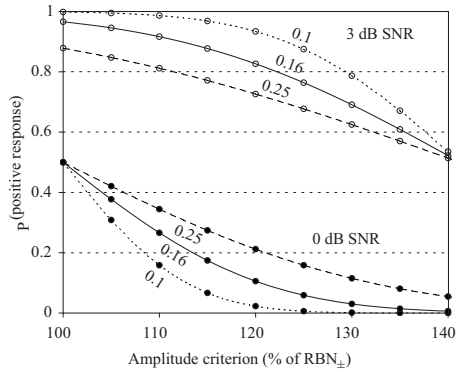


FIG. 3. Probabilities of positive judgments (false alarms at 0 dB SNR, filled circles, and hits at 3 dB SNR, open circles) as functions of varying amplitude criteria (percentage of the corresponding RBN level estimate). Curves are based on the NERs obtained from experiment 1, expressed as the ratio of no-response AEP amplitude to corresponding $RBN_{\pm}$ with an average of 1.0 ± 0.16. Predictive curves (dotted lines) are also given for datasets with the same average NERs but different SDs (values beside the curves).

Using 20% above the corresponding $RBN_{\pm}$ as the criterion to define a minimum response amplitude, the mean detection threshold is 89 ± 4.6 dB re 1 μPa (N=10) for 1000 Hz tonal burst and 92 ± 5.3 dB re 1 μPa (N=15) for noise burst.

IV. EXPERIMENT 2. THRESHOLD ESTIMATION USING STANDARDIZED AND CONTROLLED RBN AMPLITUDES

A. Introduction

Because intra- and interindividual RBN variation can be quite large, we adopted an averaging technique advocated by Don and Elberling (1996), wherein stimulus presentations were stopped after RBN amplitudes decreased below a fixed criterion, rather than after collecting a fixed number of sweeps. Given the results of experiment 1, we used the $RBN_{\pm}$ rms estimate and collected as many sweeps as needed to produce AEP averages with less than 30 nV $RBN_{\pm}$ rms. We then compared threshold estimates in individual animals at varying target RBN amplitudes.

B. Methods

Five animals were tested with both tone pips and noise bursts, as in experiment 1. The stimulus level was decreased in steps of 2.5 dB near threshold level, until several (3–5) steps below a provisionally estimated threshold. Unlike the test in experiment 1, each trial in experiment 2 was composed of flexible number of 200 sweep blocks, each containing four subblocks of 50 sweeps, with alternating stimulus polarity by subblock (100 sweeps per stimulus polarity in each block). The cumulative RBN amplitude after each additional block (200 sweeps) was monitored by calculating the $RBN_{\pm}$ of all accumulated sweeps after each block. At each stimulus level, data collection continued until the cumulative $RBN_{\pm}$ decreased below 30 nV rms (if this was reached with fewer than 2000 sweeps, additional sweeps

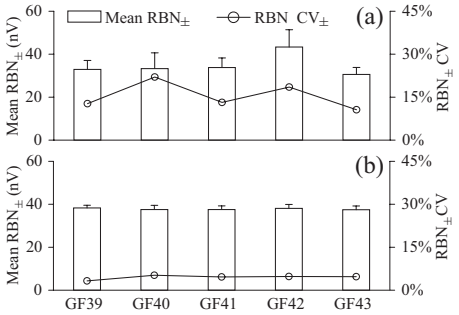


FIG. 4. Interindividual variation of $RBN_{\pm}$ obtained in experiment 2 after (a) 2000 sweeps and (b) collecting enough sweeps lowering the $RBN_{\pm}$ below the fixed target RBN amplitude (40 nV rms). Error bars are SDs.

were collected to allow for comparisons with 2000 sweep averages). In order to test the effects different RBN amplitudes might have on the threshold estimates, we extracted subaverages with $RBN_{\pm}$ amplitudes of 50, 40, and 30 nV rms from each trial. We also used the first 2000 sweeps (unfixed $RBN_{\pm}$) of this sample in our comparisons. Following the results in experiment 1, the minimal AEP amplitude containing an evoked response was defined as 120% of the corresponding $RBN_{\pm}$.

C. Results

The mean $RBN_{\pm}$ after the first 2000 sweeps (34.8 ± 4.94 nV rms, N=5) was slightly lower than that observed in experiment 1 [Fig. 4(a)]. In four of five animals, target noise amplitudes of 40 nV rms (the average amplitude in experiment 1) were obtained in fewer than 2000 sweeps (mean sweep number $= 1633 \pm 501$; N=5).

As shown in Figs. 4(a), 5, and 6(a), the noise amplitude varied considerably within and between individuals when using a fixed number (2000) of sweeps, as observed in experiment 1. The CV of mean $RBN_{\pm}$ across all individuals after 2000 sweeps in experiment 2 was 14.2%. Within individuals (or per experimental day) noise variability was similar, ranging from 10.6% to 22.0% (mean within-individual CV $= 15.4\% \pm 4.71$; N=5). As expected, however, the cumulative background noise amplitudes were less variable when averaging was stopped at a target RBN amplitude [Figs. 4(b), 5, and 6(b)], even though the decision to stop averaging is made rather coarsely, based on the cumulative noise amplitudes after each 200 sweep block. Using a target amplitude of 40 nV rms as an example, the $RBN_{\pm}$ showed a much narrower distribution compared to 2000 sweep averages (CV

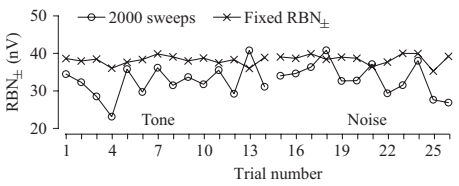


FIG. 5. Within individual fluctuation of $RBN_{\pm}$ obtained in consecutive trials from one representative animal in experiment 2 after (a) 2000 sweeps and (b) collecting enough sweeps to lower the $RBN_{\pm}$ below the fixed target RBN amplitude (40 nV rms).

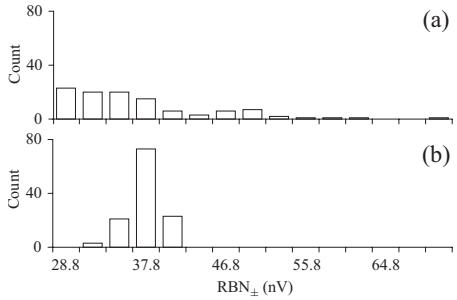


FIG. 6. Frequency distribution of $RBN_{\pm}$ obtained from all five animals used in experiment 2 after (a) 2000 sweeps and (b) collecting enough sweeps to reach the criterion $RBN_{\pm}$ amplitude (40 nV rms).

across individuals=1.0%). Variability within a subject was also greatly reduced, with a mean CV of $4.5\% \pm 0.74$ (N =5).

Figure 7 shows a detailed view of the effects of sweep number on RBN amplitudes. With increasing number of sweeps, the cumulative noise amplitude generally decreases monotonically as a power function of the number of sweeps. The AEP amplitude also declines initially as well, although it appears to asymptote or decreases its rate of attenuation after some minimum number of sweeps, if a response is contained in the average. If a criterion of the absolute amplitude of the AEP average or one relative to the RBN amplitude is used to judge the presence of an evoked response, it is clear that stopping averages at different RBN amplitudes could produce different threshold estimates. For instance, the AEP amplitude evoked by a 85.7 dB re 1 μPa tone was not 20% larger than the RBN if averaging was stopped at a target RBN amplitude of 40 nV rms, but if enough sweeps were collected to reduce the $RBN_{\pm}$ to less than 30 nV rms, then this stimulus did evoke a response 20% above RBN (Fig. 7). It is also important to note, as did Don and Elberling (1994), that the noise declines only as an approximate function of sweep number. Episodic noise sources, such as muscular contraction or external pollution can cause substantial deviations from the expected attenuation of RBN as a function of $1/\sqrt{N}$ (Fig. 7, arrows).

The mean NER obtained from this dataset at a target RBN amplitude of 40 nV rms was 1.0 ± 0.18 (the NERs were computed using AEPs evoked by stimulus levels at

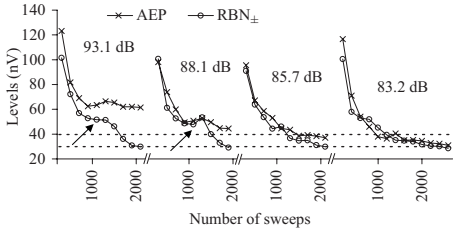


FIG. 7. Cumulative AEP and $RBN_{\pm}$ amplitudes of one representative animal in experiment 2 as functions of number of sweeps collected at several different stimulus levels (values given above the plots; re 1 μPa). Horizontal dashed line indicates two potential target RBN levels, the cumulative $RBN_{\pm}$ lower than which the collecting and averaging more sweeps can be stopped. Threshold was determined to be 88.1 dB for 40 nV rms target RBN and 85.7 dB for 30 nV rms. Arrows indicate where the cumulative RBN amplitude did not decrease at the theoretical rate of $1/\sqrt{N}$ (number of collected sweeps).

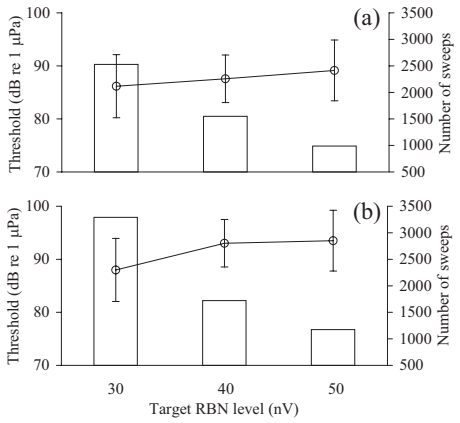


FIG. 8. The mean threshold estimates (±SD) for (a) tonal and (b) noise stimuli based on average AEPs with fixed target RBN amplitudes of 30, 40, and 50 nV rms. Histograms show the mean number of sweeps required to reach the target RBN levels.

least 5 dB lower than the estimated threshold, on average 14.0±6.79 dB below threshold). Since it was very similar to what we found and used to determine the response criterion in experiment 1, the same response criterion, 120% of the corresponding $RBN_{\pm}$, was used to estimate the threshold in this experiment.

Figure 8 shows mean detection thresholds (±SD) estimated with this method at three target RBN amplitudes. Over a range between 50 nV rms and 30 nV rms, the estimated threshold decreased by 3 dB for tonal pips (from 89±5.7 dB re 1 μPa to 86±5.9 dB re 1 μPa) and 6 dB for noise bursts (from 94±7.0 dB re 1 μPa to 88±7.1 dB re 1 μPa), respectively. Repeated measures analysis of variance found significant differences among the thresholds for tonal pips ($P=0.0049$; $P=0.1723$ for noise bursts). However, significant differences in threshold were only found between those with target RBN amplitudes of 30 nV rms and higher amplitudes for tonal stimuli (30 nV rms versus 40 nV rms: $P>0.05$; 30 nV rms vesus 50 nV rms: $P>0.01$; post hoc multiple Tukey comparison), and to decrease the $RBN_{\pm}$ from 40 nV rms to 30 nV rms, an average of 77% more sweeps (mean sweep number for 30 nV rms=2894±1025 sweeps; N=5) were required. Nonetheless, it is clear that RBN amplitude has an impact on estimated threshold and should be reported with AEP results.

V. DISCUSSION

As a method of threshold estimation, visual inspection has the clear advantage of being immediate and easy to implement. It is potentially effective, given the great human capability for visual pattern recognition and the general robustness of AEP average waveform shape. Multiple trained observers are not highly consistent, however. Both the present study and studies of trained audiologists (Arnold, 1985; Gans *et al.*, 1992; Vidler and Parker, 2004; Lv *et al.*, 2007) have reported poor agreement between observers in judging AEP threshold. In the present study, we reported that disagreement between all three observers was found for 20% of all threshold judgments, which is higher than what Lv *et al.* (2007) reported for three experienced audiologists (one out of twelve judgments in their Fig. 4). Cohen's Kappa values, which measure observer agreements, were very similar in Arnold (1985), Lv *et al.* (2007), and the present study and were all below 0.9, a standard minimum criterion for independent observer reliability. This lack of reliability makes visual inspection particularly unsuited for hypothesis testing using AEP unless double-blind data collection and examination is carried out, which is generally not practical. Moreover, the need to have the AEP examined by multiple experienced observers, which might be necessary for reducing the subjective bias, is not always satisfied in AEP studies on animals. Nor is there a clear method for resolving observer discrepancies, other than averaging. Given these reasons, an objective method of threshold determination is highly desirable.

The determination of detection threshold depends on the quality of recorded average responses, and is therefore very likely affected by changes in background noise amplitude. Methods that depend on response repeatability (e.g., visual inspection or correlation analysis) can be affected by the distorting effects of noise in near threshold AEP waveforms (Elberling and Don, 1984). Methods that depend on response amplitude (e.g., Supin *et al.*, 2001) must also consider noise amplitude because the AEP amplitude reflects the sum of the RBN and the evoked potential together, rather than the size of the evoked response alone. Moreover, even methods based on the relationship between AEP and RBN amplitude, like the one proposed herein, may still differ in sensitivity if different target RBN amplitudes are sought (or if RBN is not controlled). For example, we found a significant effect of reducing RBN amplitudes on threshold estimation when using tonal stimuli (Fig. 8). In our samples, and in similar reports in the literature (Elberling and Don, 1984), RBN amplitude can vary greatly after equal numbers of sweeps, and it is also possible that experimental treatments (such as drug-effects) could alter background physiological noise amplitudes. Standardized noise amplitudes would therefore minimize one confounding source of variability in AEP studies and should be used in all studies, regardless of threshold estimation technique. We therefore suggest that the presence of a response should be judged in relation to background noise amplitude, and that in hypothesis testing all comparisons should be made at equal or similar RBN amplitudes.

One advantage of the present method over subjective methods is that it allows for estimation of the accuracy of the technique based on the particular combinations of empirical noise reliability estimates and amplitude criteria (e.g., Fig. 3), as do other objective methods (Elberling and Don, 1987). Elberling and Don (1984) have advocated a parametric statistical approach using the F-distribution to compare the ratio (F_{SP}) of EP variance (the square of AEP rms) to noise variance (the square of RBN_{SP}). After estimating the degrees of freedom in an AEP average, the F_{SP} can be compared to standard critical values of the F-distribution. The variance in single point amplitudes is a robust estimate of RBN variance and is thought to have degrees of freedom equal to the number of samples (i.e., sweeps). The AEP rms, however, is a different kind of time average whose degrees of freedom are not equal to the number of sweeps, but rather are limited by

the frequency content present in the averaged time-window. Elberling and Don (1984) have provided estimated degrees of freedom for AEPs that range from 8 to 22. Unfortunately, much of our single-point values were not normally distributed and our attempts to fit an F-distribution to no-signal AEP rms^2/RBN_{SP}^2 ratios were not effective for estimating the degrees of freedom of our AEP rms. Instead we used an estimate of noise reliability that was amenable to a detection theory analysis rather than analysis by F-distribution critical values. Although the $RBN_{\pm}$ has lower degrees of freedom than RBN_{SP}, it is equal in degree of freedom to the AEP rms estimate.

Error rates (misses and false alarms) can be predicted using our measure of noise estimate reliability (Fig. 3). Different values used as a minimum response amplitude (120% in present study) would result in different error rates (for both false alarms at 0 dB SNR and misses at higher SNRs) if the reliability of the noise estimate is unchanged, as shown in Fig. 3. Further, if the reliability of the RBN estimate differs, error rates will also change for any specific criterion value. Nevertheless, as long as the NER is empirically measured and a minimum response criterion is chosen, hit and false alarm rates for any specific SNR of AEP averages can be easily predicted. Moreover, our approach could also be used to estimate the expected detectability [in standard signal detection measures such as d' (Green and Swets, 1988)] of a given SNR or the lowest SNR that would result in a specific d' value. For instance, if $NER=1.0\pm0.16$ and a criterion of 120% of $RBN_{\pm}$ is used to judge the presence of a response, d' values greater than 1 are expected to occur whenever the AEP SNR is greater than 1.24 dB.

The choices of target RBN amplitudes and the criteria depend on the experimental preparation and the amount of environmental noise present in the laboratory, both of which will influence the time required (number of sweeps) to reach the target noise amplitudes. Although a lower target RBN amplitude might provide a more sensitive threshold estimate (Figs. 7 and 8), the number of sweeps required will also increase and may prolong data collection beyond practical limits (Fig. 8). Therefore, the choice of the target RBN amplitude can only be made by a tradeoff between the desired quality of the AEP, i.e., the RBN amplitude, and the affordable time expense.

The choice of noise measurement may also affect the sensitivity of the threshold estimation (e.g., Cebulla *et al.* 2000; Stürzebecher *et al.*, 2001). In the present study, measures of RBN_{SP} were not always normally distributed, which did not match the assumptions required to use this measure in the statistical fashion proposed by Elberling and Don (1984). It is difficult to explain the difference in our data from those presented by Elberling and Don (1984). One potential explanation may be that the noise distribution was skewed by episodic noise from either the subject (e.g., muscle activity) or nearby electronics. However, we did not find obvious correlation between the non-normal blocks and the RBN amplitude or the number of sweeps rejected within the collection of the data, and few of the non-normal single-point blocks were significantly skewed. In any case, we found that $RBN_{\pm}$ was statistically indistinguishable from no-response AEP amplitudes while RBN_{SP} was significantly higher than direct measures of RBN amplitude in no-response averages. Other studies (in humans) have not reported such difficulties, and some have indicated that RBN_{SP} may be a more sensitive measure than $RBN_{\pm}$ (Don *et al.*, 1984; Stürzebecher *et al.* 2001).

We also prefer the $RBN_{\pm}$ because it can be easily measured by most commercial software packages with few modifications; the procedure with a controlled RBN amplitude can be simply implemented by postprocessing of blocks of small numbers of sweeps and manually deciding to continue or stop collecting sweeps. It can be performed either with or without a true on-line analysis or the storage of single sweeps. In any case, the procedures advocated above, including controlling RBN and using an AEP:RBN ratio to determine threshold, may be applied using any potential method of noise amplitude estimation, including F_{SP}. Further, we advocate an internal measure of RBN reliability that can be used in a signal detection analysis. Specific hit and false alarm rates can be achieved by adjusting the exact AEP:RBN ratio criterion based on each noise estimate, much like defining α in an F-ratio with a given degree of freedom (as with F_{SP}: Elberling and Don, 1987).

The threshold values reported herein are higher than most behavioral thresholds (see Fay, 1988 for review; but also see Wolski *et al.*, 2003 for a discussion of differences between behavioral and AEP thresholds in the same subjects) and some AEP thresholds (e.g., Kenyon *et al.*, 1998, 65 dB re 1 μPa), but are similar to those published by Lu and Tomchick (2002, ~80 dB re 1 μPa) and are consistent across multiple groups of subjects in this laboratory (e.g., Cordova and Braun, 2007, 83 dB re 1 μPa). Similar large differences between laboratories have been greatly discussed (Popper *et al.*, 1973; Hawkins, 1981; Fay, 1988), and may reflect large differences in the acoustics of underwater sound propagation in different tanks (Parvulescu, 1967; Akamatsu *et al.*, 2002). Stimuli with wavelengths larger than the arena (the wavelength of 1000 Hz underwater is 1.5 m) result in hydrodynamic fields with great (and unpredictable) inhomogeneities in the particle velocities in the sound field. The size, shape, and material properties of the arena greatly influence the pressure-to-velocity ratio for any given sound source, thus the position of the fish and speaker will change the true stimulus to the subject. A threshold measured only in pressure will not guarantee equivalent stimuli across different positions within the arena. Laboratories also differ in the method of holding the subject, which could easily affect the response of displacement sensitive otolithic organs. These factors greatly complicate the comparison of threshold values across laboratories, and future studies should examine the role of tank acoustics and AEP methodologies for measuring fish auditory thresholds.

It is less appreciated, however, that objective measures of AEP threshold are even more important for comparisons of hearing function *within the same laboratory*. Subjective threshold determination techniques are not consistent across observers, which raises the potential for subjective bias. Since physiological experiments are generally not conducted in a double-blind fashion, some objective thresholding tech-

nique is essential for statistical hypothesis testing. We suggest that controlled and clearly reported RBN amplitudes will effectively eliminate potential bias and intralaboratory variability in AEP audiometry.

VI. SUMMARY AND CONCLUSIONS

In this report, we evaluated the interobserver reliability of subjective response determination in goldfish AEP audiometry using tone and noise stimuli. We found, as has been reported in several similar studies of clinical audiologists, that visual inspection and subjective evaluation is only modestly reliable at best. Among the three experienced observers in this study, complete and partial agreement on threshold level was found in 80% of their judgments. One potential explanation for this disagreement is that low signal-to-noise ratios in near-threshold AEPs make these judgments difficult. This phenomenon raises the added problem that if noise amplitudes vary, either between animals or experimental treatment groups, the difficulty, and hence, the reliability of subjective threshold determination techniques will differ between groups. This represents a serious pitfall for the use of AEPs in testing hypotheses about auditory function, for example after sonic or toxic insult. To develop a more objective response determination technique and measure the potential for noise variability, we evaluated two measures of residual noise in AEP averages. We found that while both measures provided relatively similar estimates of noise amplitude, RBN_{SP} did not meet the normality assumption required for its use (Elberling and Don, 1984) in our data. Noise estimates as $RBN_{\pm}$ were indistinguishable from direct measurements of AEP physiological noise from no-response AEP recordings and has the added advantage of being easy to calculate with nearly any commercial evoked-potential software package. We further described the statistical structure of the ratio between $RBN_{\pm}$ and direct measures of AEP noise to estimate the reliability of our noise estimate, and to estimate the hit and false alarm rates of objective criteria for specific minimal response AEP SNRs. We conclude that objective measures of threshold determination with RBN amplitude control are relatively simple to apply and should be used in hypothesis testing using AEP audiometry. At a minimum, future AEP studies should include the verification of equal noise amplitudes in experimental treatment groups.

ACKNOWLEDGMENTS

The authors thank Dr. Manuel Don for his encouragement and advice during the development of this technique. Micah Cordova acted as our third subjective observer and they thank him for his contribution. Financial support for this research was provided by NIH awards 1 S06 GM60654 and 1 R03 MH067808 to C.B.B. The National Center for Research Resources of the NIH supports research infrastructure at Hunter College through the Research Centers in Minority Institutions (RCMI) award RR-03037.

Akamatsu, T., Okumura, T., Novarini, N., and Yan, H. Y. (**2002**). "Empirical refinements applicable to the recording of fish sounds in small tanks," J. Acoust. Soc. Am. **112**, 3073–3082.

Arnold, S. A. (**1985**). "Objective versus visual detection of the auditory brain stem response," Ear Hear. **6**, 144–150.

Boettcher, F. A. (**2002**). "Susceptibility to acoustic trauma in young and aged gerbils," J. Acoust. Soc. Am. **112**, 2948–2955.

Brittan-Powell, E. F., Dooling, R. J., and Gleich, O. (**2002**). "Auditory brainstem responses in adult budgerigars (*Melopsittacus undulatus*)," J. Acoust. Soc. Am. **112**, 999–1008.

Burkard, R. (**1984**). "Sound pressure level measurement and spectral analysis of brief acoustic transients," Electroencephalogr. Clin. Neurophysiol. **57**, 83–91.

Cebulla, M., Stürzebecher, E., and Wernecke, K.-D. (**2000**). "Objective detection of auditory brainstem potentials." Scand. Audiol. **29**, 44–51.

Cohen, J. (**1960**). "A coefficient of agreement for nominal scales," Educ. Psychol. Meas. **20**, 37–46.

Cordova, M. S., and Braun, C. B. (**2007**). "The use of anesthesia during evoked potential audiometry in goldfish (*Carassius auratus*)," Brain Res. **1153**, 78–83.

Don, M., and Elberling, C. (**1994**). "Evaluating residual background noise in human auditory brain-stem responses," J. Acoust. Soc. Am. **96**, 2746–2757.

Don, M., and Elberling, C. (**1996**). "Use of quantitative measures of auditory brain-stem response peak amplitude and residual background noise in the decision to stop averaging," J. Acoust. Soc. Am. **99**, 491–499.

Don, M., Elberling, C., and Waring, M. (**1984**). "Objective detection of averaged auditory brainstem responses," Scand. Audiol. **13**, 219–228.

Elberling, C. (**1979**). "Auditory electrophysiology. The use of templates and cross correlation functions in the analysis of brain stem potentials," Scand. Audiol. **8**, 187–190.

Elberling, C., and Don, M. (**1984**). "Quality estimation of averaged auditory brainstem responses," Scand. Audiol. **13**, 187–197.

Elberling, C., and Don, M. (**1987**). "Threshold characteristics of the human auditory brain stem response," J. Acoust. Soc. Am. **81**, 115–121.

Elberling, C., and Wahlgreen, O. (**1985**). "Estimation of auditory brainstem response, ABR, by means of Bayseian inference," Scand. Audiol. **14**, 89–96.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Associates, Winnetka).

Gans, D., Del Zotto, D., and Gans, K. D. (**1992**). "Bias in scoring auditory brainstem responses," Br. J. Audiol. **26**, 363–368.

Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos).

Hawkins, A. D. (**1981**). "The hearing abilities of fish," in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer, New York), pp. 109–133.

Higgs, D. M., Brittan-Powell, E. F., Soares, D., Souza, M. J., Carr, C. E., Dooling, R. J., and Popper, A. N. (**2002**). "Amphibious auditory responses of the American alligator (*Alligator mississipiensis*)," J. Comp. Physiol., A **188**, 217–223.

Higgs, D. M., Rollo, A. K., Souza, M. J., and Popper, A. N. (**2003**). "Development of form and function in peripheral auditory structures of the zebrafish (*Danio rerio*)," J. Acoust. Soc. Am. **113**, 1145–1154.

Hyde, M., Sininger, Y. S., and Don, M. (**1998**). "Objective detection and analysis of auditory brainstem response: an historical perspective," Semin. Hear. **19**, 97–113.

Katbamna, B., Brown, J. A., Collard, M., and Ide, C. F. (**2006**). "Auditory brainstem responses to airborne sounds in the aquatic frog Xenopus laevis: Correlation with middle ear characteristics," J. Comp. Physiol., A **192**, 381–387.

Kenyon, T. N., Ladich, F., and Yan, H. Y. (**1998**). "A comparative study of hearing ability in fishes: the auditory brainstem response approach," J. Comp. Physiol., A **182**, 307–318.

Le Prell, C. G., Yagi, M., Kawamoto, K., Beyer, L. A., Atkin, G., Raphael, Y., Dolan, D. F., Bledsoe, S. C. J., and Moody, D. B. (**2004**). "Chronic excitotoxicity in the guinea pig cochlea induces temporary functional deficits without disrupting otoacoustic emissions," J. Acoust. Soc. Am. **116**, 1044–1056.

Lu, Z., and Tomchik, S. M. (**2002**). "Effects of a red-tide toxin on fish hearing," J. Comp. Physiol., A **188**, 807–813.

Lv, J., Simpson, D. M., and Bell, S. L. (**2007**). "Objective detection of evoked potentials using a bootstrap technique," Med. Eng. Phys. **29**, 191–198.

Mann, D. A., Higgs, D. M., Tavolga, W. N., Souza, M. J., and Popper, A. N. (**2001**). "Ultrasound detection by clupeiform fishes," J. Acoust. Soc. Am. **109**, 3048–3054.

May, B. J., Prosen, C. A., Weiss, D., and Vetter, D. (**2002**). "Behavioral

investigation of some possible effects of the central olivocochlear pathways in transgenic mice," Hear. Res. **171**, 142–157.
Mitchell, C., Kempton, J. B., Creedon, T., and Trune, D. (**1996**). "Rapid acquisition of auditory brainstem responses with multiple frequency and intensity tone-bursts.," Hear. Res. **99**, 38–46.
Nickerson, R. S. (**1998**). "Confirmation bias: A ubiquitous phenomenon in many guises," Rev. Gen. Psychol. **2**, 175–220.
Ozdamar, O., Delgado, R. E., Eilers, R. E., and Urbano, R. C. (**1994**). "Automated electrophysiologic hearing testing using a threshold-seeking algorithm," J. Am. Acad. Audiol **5**, 77–88.
Parvulescu, A. (**1967**). "The acoustics of small tanks," in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, Oxford), pp. 7–13.
Popper, A. N., Chan, A. T. H., and Clarke, N. L. (**1973**). "An evaluation of methods for behavioral investigations of teleost audition," Behav. Res. Methods Instrum. **5**, 470–472.
Popper, A. N., Smith, M. E., Cott, P. A., Hanna, B. W., MacGillivray, A. O., Austin, M. E., and Mann, D. A. (**2005**). "Effects of exposure to seismic airgun use on hearing of three fish species," J. Acoust. Soc. Am. **117**, 3958–3971.
Porter, C. A., Martin, G. K., Stagner, B. B., and Lonsbury-Martin, B. L. (**2006**). "Distortion-product otoacoustic emission suppression growth in normal and noise-exposed rabbits," J. Acoust. Soc. Am. **120**, 884–900.
Ramcharitar, J., and Popper, A. N. (**2004**). "Masked auditory thresholds in sciaenid fishes: A comparative study," J. Acoust. Soc. Am. **116**, 1687–1691.
Ramcharitar, J. U., Higgs, D. M., and Popper, A. N. (**2006**). "Audition in sciaenid fishes with different swim bladder-inner ear configurations," J. Acoust. Soc. Am. **119**, 439–443.
Schimmel, H. (**1967**). "The (±) reference: accuracy of estimated mean components in average response studies," Science **157**, 92–94.
Song, L., McGee, J., and Walsh, E. J. (**2006**). "Frequency- and level-dependent changes in auditory brainstem responses (ABRs) in developing mice," J. Acoust. Soc. Am. **119**, 2242–2257.
Stürzebecher, E., Cebulla, M., and Wernecke, K.-D., (**2001**). "Objective detection of transiently evoked otoacoustic emissions," Scand. Audiol. **29**, 78–88.
Supin, A. Y., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer, New York).
Vidler, M., and Parker, D. (**2004**). "Auditory brainstem response threshold estimation: Subjective threshold estimation by experienced clinicians in a computer simulation of the clinical test," Int. J. Audiol. **43**, 417–429.
Wolski, L. F., Anderson, R. C., Bowles, A. E., and Yochem, P. K. (**2003**). "Measuring hearing in the harbor seal (*Phoca vitulina*): Comparison of behavioral and auditory brainstem response techniques.," J. Acoust. Soc. Am. **113**, 629–637.
Wong, P. K. H., and Bickford, R. G. (**1980**). "Brain stem auditory evoked potentials: The use of noise estimate," Electroencephalogr. Clin. Neurophysiol. **50**, 25–34.
Yan, H. Y. (**1998**). "Auditory role of the suprabranchial chamber in gourami fish," J. Comp. Physiol., A **183**, 325–333.

The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms

Nicole Marrone,[a] Christine R. Mason, and Gerald Kidd, Jr.
Department of Speech, Language, and Hearing Sciences and the Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 3 August 2007; revised 30 July 2008; accepted 31 July 2008)

This study investigated the interaction between hearing loss, reverberation, and age on the benefit of spatially separating multiple masking talkers from a target talker. Four listener groups were tested based on hearing status and age. On every trial listeners heard three different sentences spoken simultaneously by different female talkers. Listeners reported keywords from the target sentence, which was presented at a fixed and known location. Maskers were colocated with the target or presented from spatially separated and symmetrically placed loudspeakers, creating a situation with no simple "better-ear." Reverberation was also varied. The target-to-masker ratio at threshold for identification of the fixed-level target was measured by adapting the level of the maskers. On average, listeners with hearing loss showed less spatial release from masking than normal-hearing listeners. Age was a significant factor although small differences in hearing sensitivity across age groups may have contributed to this effect. Spatial release was reduced in the more reverberant room condition but in most cases a significant advantage remained. These results provide evidence for a large benefit of spatial separation in a multitalker situation that is likely due to perceptual factors. However, this benefit is significantly reduced by both hearing loss and reverberation.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2980441]

PACS number(s): 43.66.Dc, 43.66.Lj, 43.66.Pn, 43.66.Sr [RLF] Pages: 3064–3075

I. INTRODUCTION

It is well documented that spatial separation of one or more interfering sound sources from a target source may result in significant release from masking in normal-hearing listeners [see reviews by Yost (1997), Bronkhorst (2000), Ebata (2003), and Colburn *et al.* (2006)]. In general, there have been fewer psychophysical studies of spatial release from masking in listeners with hearing loss, especially when the interfering sources are competing talkers in real rooms. However, that type of situation is known to be among the more challenging communication environments for listeners with hearing loss and as such it is important to fully understand the factors underlying their difficulties. Gatehouse and Noble (2004) designed the speech, spatial, and qualities of hearing scale (SSQ) to assess the perceived difficulty of listening in complex and dynamic auditory environments. They found that the items related to aspects of auditory selective attention and switching attention (e.g., having a conversation with one person when many people are talking, following the TV while talking to someone, etc.) were those most highly correlated with experiencing handicap as a result of hearing loss. Thus, it would seem that there is a compelling rationale for studying selective listening in listeners with hearing loss in complex sound environments.

Traditionally, spatial release from masking has been thought to be a consequence of relatively low-level (i.e., peripheral) factors, including the simple acoustic effect of "better-ear listening" [attending to the ear with the higher target-to-masker (T/M) ratio] and binaural analysis (leading to "bottom-up" within-channel improvements in T/M, or "masking level differences"). Models of this process based on interaural level differences (ILDs) and interaural time differences (ITDs) are generally successful in predicting the improvement in the intelligibility of target speech masked by noise when the two sources are spatially separated (e.g., Zurek, 1993; Bronkhorst, 2000). Yet when the listening situation is more complex and the maskers are highly distracting similar sounds such as one or more competing streams of speech (often producing large amounts of centrally based "informational masking," cf. Kidd *et al.*, 2008), a large release from masking can be observed, which is not accounted for by lower-level peripheral factors and is not successfully predicted by such models. The results of several recent studies using more complex stimuli and challenging listening environments have shown that spatial release from masking is a complicated phenomenon comprising both lower-level *and* higher-level processes (e.g., Kidd *et al.*, 1998; Freyman *et al.*, 1999; Hawley *et al.*, 1999; Noble and Perrett, 2002; Arbogast *et al.*, 2002; Hawley *et al.*, 2004; Best *et al.*, 2006; Marrone *et al.*, 2008). Currently, very little is known about the role of nonperipheral factors in complex listening situations in listeners with hearing loss. Thus, the goal of the current study was to examine higher-level factors (e.g., selective attention) in spatial release from masking and determine the extent to which the benefit of spatially separating competing talkers is affected by sensorineural hearing loss and age.

The use of a single masking source or the asymmetric positioning of multiple masker sources to create spatial release may make the separation of the contributions of lower-

[a]Author to whom correspondence should be addressed. Electronic mail: n-marrone@northwestern.edu

 © 2008 Acoustical Society of America

level and higher-level processes difficult to ascertain because of the resulting opportunity for the listener to benefit from the improvement in T/M in one ear due to acoustic head shadow. Consequently, the maskers in the current study were placed symmetrically around the target in order to diminish or eliminate any potential for a better-ear advantage. Furthermore, the stimuli and methods used are known to produce large amounts of informational masking, which in young normal-hearing listeners may be substantially reduced by spatially separating the sound sources (Arbogast *et al.*, 2002; Brungart and Simpson, 2002; Shinn-Cunningham *et al.*, 2005; Marrone *et al.*, 2008). The main finding of the study of Marrone *et al.* (2008), which employed the same stimuli and similar methods to those used here, was a tuned response in azimuth. More specifically, the T/M for speech identification threshold progressively decreased as the two symmetric speech maskers were varied in azimuth from 0° (colocated with target) to 45°. The benefit of spatial separation between the target and maskers had an asymptotic maximum of 12 dB on average. Importantly, a control condition in which listeners had one ear occluded by an earplug and earmuff revealed that there was no release from masking with increasing spatial separation of the maskers when the listener could only listen monaurally. Although binaural cues are thus fundamentally important to the task, within-channel binaural analysis was not considered to be a significant factor in producing the spatial release observed in that study because it was concluded that (a) relatively little within-channel energetic masking was present (cf. Brungart *et al.*, 2006), suggesting that masking level differences (MLDs) would be minimal; (b) the amount of release predicted by the Bronkhorst (2000) model (for speech in noise, approximately 2 dB) was much smaller than that which was observed, also consistent with the notion that the task was dominated by informational masking; and (c) a noise control condition (producing primarily energetic masking) resulted in very little spatial release (about 1–2 dB), in agreement with that model's prediction.

Marrone *et al.* (2008) speculated that segregating the three talkers was a very difficult task when the talkers were colocated primarily because of the high degree of target-masker similarity (all talkers were female) as well as target talker uncertainty (the target talker was chosen randomly from among them on each trial). In this difficult listening situation, identification of the target speech was only reliably possible when the level of the target was higher than that of the maskers. In support of that idea, they simply reversed the masker speech, thereby decreasing the target-masker similarity (and the opportunity for target-masker confusion), and improved the threshold T/M significantly (by about 12 dB). Because of the possibility that energetic masking could have differed for forward and reversed speech maskers (due to overall envelope differences or because the target and masker words are aligned differently in the two conditions), Marrone *et al.* (2008) also tested a control condition using speech-shaped speech-envelope-modulated noise as a masker. The speech envelope was taken from either forward or reversed speech maskers and the amount of masking produced in the two conditions was compared. The results of that control condition supported the conclusion that the difference for forward versus backward speech maskers was not due to differences in energetic masking. Furthermore, a release from masking for the forward speech of nearly 8 dB was observed despite presentation in a more reverberant listening condition. For the reasons listed above and others considered in that article, Marrone *et al.* (2008) concluded that the spatial release observed was due primarily to improved speech stream segregation and stronger focus of attention along the spatial dimension. Although this process is still not fully understood and clearly depends on interaural differences, Marrone *et al.* (2008) hypothesized that the presence of a high degree of informational masking was necessary (although perhaps not sufficient) to observe large effects with symmetric speech maskers.

Several previous studies have examined the utility of spatial cues in overcoming masking in listeners with hearing loss. Generally, these studies have used noise or multitalker babble as maskers and have found smaller spatial release in listeners with hearing loss than with normal hearing (Duquesnoy, 1983; Gelfand *et al.*, 1988; Bronkhorst and Plomp, 1989, 1992; Helfer, 1992; Ter-Horst *et al.*, 1993; Peissig and Kollmeier, 1997; Dubno *et al.*, 2002). Noise maskers and multitalker babble may be dominated by energetic masking (cf. Freyman *et al.*, 2004) and so the extent to which these studies are informative about higher-level selective listening processes is unclear. In fact, one frequently cited hypothesis regarding the reduced spatial release from masking in listeners with hearing loss has to do with the reduced opportunity to take advantage of an acoustically better ear. Dubno *et al.* (2002) found that the benefit of spatially separating speech-shaped noise from a target talker was differentially affected by low-pass and high-pass filtering in younger and older adults with normal hearing and older adults with hearing loss. Whereas the two normal-hearing groups demonstrated significant benefit from source separation in both filter conditions, the older adults with hearing loss received little benefit when the speech and noise were high-pass filtered (presumably providing primarily ILD cues) but showed a small benefit from the low-pass speech and noise (probably related primarily to ITDs). This leads to the question of whether the benefit from spatial cues would be similar for listeners with normal hearing and those with hearing loss if no overall better-ear advantage (in contrast to time-varying or momentary, discussed more fully below) were available in the stimulus.

A few of the studies of spatial release from masking in listeners with hearing loss have included competing speech as a masker (Duquesnoy, 1983; Peissig and Kollmeier, 1997; Arbogast *et al.*, 2005; Helfer and Freyman, 2008). These studies have also mostly used single source/asymmetric masker placements and generally have found a decreased benefit of spatial separation in listeners with hearing loss. For example, Duquesnoy (1983) found that the benefit of spatially separating a female talker from a male competing talker by 90° was 6.7 dB on average for young normal-hearing listeners while it was 4.5 dB for older listeners with hearing loss. However, it was not clear whether this difference was due to hearing loss, the difference in age between

the listener groups, or an interaction between these two factors. Furthermore, the relatively small spatial release may have indicated that the different-sex talkers were fairly easy to segregate and produced little informational masking, so the further benefit of providing spatial cues was limited. Peissig and Kollmeier (1997) also tested speech reception in the presence of multiple speech maskers but under conditions of relatively low uncertainty, using known target sentences with continuous interference in a subjective intelligibility procedure to estimate speech threshold with three masking sources. The maximum benefit of spatial separation in this condition was 5 dB on average for the listeners with normal hearing while it ranged from 0.6 to 3.8 dB for the listeners with hearing loss. Arbogast *et al.* (2005) used a form of cochlear implant simulation processing (e.g., Shannon *et al.*, 1995) to filter the target and masker speech into different frequency bands in order to limit energetic masking. They found that the mean spatial release from masking for a 90° horizontal separation between two male talkers was about 15 dB for the listeners with normal hearing and about 10 dB for age-matched listeners with hearing loss. These results were interpreted as evidence that both types of listeners can benefit from spatial separation between a target talker and a speech masker. However, the amount of spatial release was less in the listeners with hearing loss, possibly because of greater energetic masking due to wider auditory filters. Evidence in support of this idea came from an off-frequency noise masking control condition in which the listeners with hearing loss showed significantly greater masking than those with normal hearing. Thus, the asymmetric single masker placement allowed for better-ear cues and the presumably poorer frequency selectivity in the impaired ears may have increased the overlap of excitation between the sharply filtered bands increasing the ratio of energetic to informational masking.

In most of the studies above in which normal-hearing and hearing-impaired listeners were compared, the hearing-impaired listeners were older than their normal-hearing counterparts. There have been a number of studies suggesting that age is a factor in either selective listening tasks (e.g., Helfer and Freyman, 2008) or divided listening tasks (e.g., Humes *et al.*, 2006) separate from the consequences of hearing loss. In a recent study by Dubno *et al.* (2008) it was concluded that older listeners may have deficits in binaural processing for sentence recognition (measured in noise but without better-ear cues by use of symmetric masker placement) even though no age-related differences were found for the tasks involving simple detection and recognition. Furthermore, there are indications that advanced age affects other higher-level processes in hearing such as auditory memory (e.g., Pichora-Fuller *et al.*, 1995). Because the tasks and listening situations we are using in these experiments are explicitly intended to tap higher-level processes such as selective attention and memory, age is an important factor that is incorporated into the current experimental design. In addition, because the majority of available listeners with hearing loss tend to be older, it is important to control for the factor of age separately. Thus, we are attempting to determine whether older listeners—either with or without hearing loss—exhibit greater difficulty in a complex multitalker environment than matched younger listeners.

The current study applies the methods and stimuli used by Marrone *et al.* (2008) to examine the benefit of spatial separation between multiple competing talkers in younger and older listeners with and without sensorineural hearing loss. The goal is to emphasize the contributions of higher-level processes in spatial release from masking in part through the use of a procedure that severely restricts better-ear acoustic cues and that normally causes large amounts of informational masking. Furthermore, in an attempt to approximate more realistic listening environments where listeners with hearing loss report experiencing the greatest difficulty, masked speech identification is tested under two levels of reverberation.

II. METHODS

A. Listeners

Four listener groups were recruited based on hearing status (normal hearing/hearing loss) and age (younger/older). All listeners had symmetric hearing thresholds ($\leq$10 dB difference between ears at any audiometric test frequency) and no significant air-bone gap ($<$10 dB difference between air and bone conduction thresholds at any audiometric test frequency in one ear). The normal-hearing groups had audiometric thresholds $\leq$25 dB hearing level (HL) at octave frequencies from 250 to 4000 Hz. The listeners with hearing loss had mild-to-moderately severe flat or sloping sensorineural hearing loss. All of the listeners with hearing loss wore aids bilaterally but were tested unaided in this experiment. All of the listeners spoke American English as their primary language and were paid for their participation.

The group of ten younger listeners with hearing loss (YHL) had an age range of 19–42 years with a mean age of 28.5 years. They were age-matched with ten younger listeners with normal hearing (YNH), age 22–43 years with a mean age of 31.4 years. The older hearing loss group (OHL) had ten listeners ranging in age from 57 to 80 years with a mean age of 69.9 years. The older normal-hearing (ONH) group had a slightly smaller age range, 59–74 years, but the mean age (66.2 years) was similar to the OHL group. Figure 1 displays the mean audiograms for the four listener groups. Mean thresholds in dB HL across audiometric frequencies from 250 to 8000 Hz are shown for each listener group for both ears. The lowest thresholds across frequency were for the YNH group. The ONH group had slightly higher thresholds at all test frequencies. For the listeners with hearing loss, mean thresholds for the OHL group were within 10 dB HL of those for the YHL group from 250 to 4000 Hz.

B. Stimuli

The stimuli were recordings of the four female talkers from the coordinate response measure (CRM) corpus (Bolia *et al.*, 2000), in which every sentence has the following structure: "Ready [callsign] go to [color] [number] now." The target sentence always had the callsign "Baron," although the talker varied from trial to trial. The maskers were

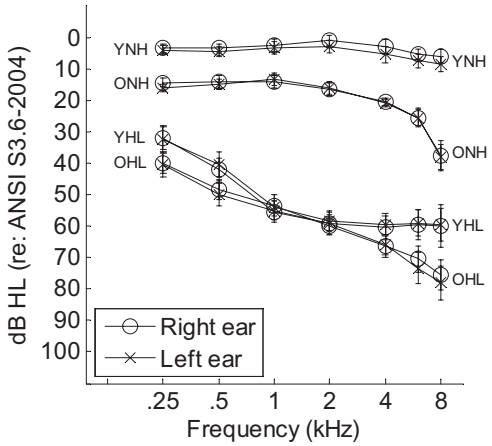


FIG. 1. Mean thresholds in dB HL for each listener group for right and left ears (circle and cross symbols, respectively) plotted as a function of audiometric frequency from 250 to 8000 Hz. The error bars represent ±1 standard error of the mean.

also CRM sentences spoken by two different talkers from the corpus. The masker talkers, callsigns, colors, and numbers were different from the target and from each other and varied from trial to trial.

C. Room conditions

The experiment was conducted in the Soundfield Laboratory at Boston University with stimuli presented through loudspeakers (Acoustic Research 215 PS). The room is a large single-walled Industrial Acoustics Company (IAC) sound booth (12 ft, 4 in long; 13 ft wide; and 7 ft, 6 in. high) in which the reverberation characteristics can be changed by covering all surfaces (ceiling, floor, walls, and door) with custom fit panels of different acoustic reflectivities. All experimental conditions were completed in two sessions, one for each of two room conditions. Session order was counterbalanced across listeners. In the low-reverberation condition, the surfaces were left uncovered (referred to as the "BARE" room condition). This room configuration was that of standard IAC booths: The ceiling and walls had a perforated metal surface and the floor was carpeted. In the second, more reverberant condition, all surfaces were covered with Plexiglas® panels (the "PLEX" room condition). Acoustic measurements of these room configurations were reported by Kidd *et al.* (2005), and included impulse responses, modulation transfer functions, cross-correlation functions, and ILD measurements. There are differences between the two room conditions for each of these measurements corresponding to the increase in reverberation due to the reflective panels in the PLEX room. For example, there was a decrease in the direct-to-reverberant energy ratio of more than 7 dB (from 6.3 to −0.9 dB for BARE versus PLEX, respectively) as measured at the approximate position of the listener's head. Also, reverberation times increased by a factor of 4 from around 0.06 s in BARE to just over 0.25 s in PLEX.[1]

D. Procedures

1. General

The experimental setup and procedures used were similar to those in Marrone *et al.*, (2008). As mentioned above, there were two sessions each lasting approximately 2 h that were typically scheduled within a few days of each other. Listeners were seated in the sound booth with an array of seven loudspeakers arranged in a semicircle in the horizontal plane; however, stimuli were only presented from the three loudspeakers positioned in front of the listener (0°) and to either side (±90°) depending on the condition. The faces of the loudspeakers were 5 ft from the approximate location of the center of the listener's head when seated. Listeners used a handheld keypad with a liquid crystal display (Q-term II) to enter their responses and receive feedback on each trial. The word "listen" appeared on the display at the beginning of each trial. After stimulus presentation, listeners responded to the prompts, "Color [B R W G]?" and "Number [1–8]?" on the keypad. The feedback consisted of a message indicating whether the color and number reported were correct and what the target color and number had been for that trial (for example, "Correct, it was blue four").

The computer and Tucker-Davis Technologies hardware used to control the experiment and present the stimuli were located outside the booth. Stimuli were played at a 40 kHz rate via a 16 bit, eight-channel digital-to-analog converter, low-pass filtered at 20 kHz, and attenuated. The target was routed through a programmable switch. On trials when the maskers were spatially separated from the target (±90° spatial separation), each sentence was routed through separate digital-to-analog converter channels, filters, attenuators, power amplifiers (Tascam), and played through separate loudspeakers. On trials when the target and maskers were colocated (0° spatial separation), the two masker sentences were first digitally added. The combined two-talker masker and the target were then routed through separate digital-to-analog converter channels, filters, and attenuators before being combined in a mixer, passed through a power amplifier, and sent to the loudspeaker.

Prior to each session, the system was calibrated so that the loudspeakers were correctly positioned and the output level measured with a Brüel & Kjær microphone suspended in that position for a given input was verified and the same from each loudspeaker. For a broadband noise of the same level at the input to the loudspeakers measured at the position of the listener's head, the overall sound pressure level (SPL) was approximately 3 dB higher in the PLEX room than in the BARE room. This correction was made when the results are reported in dB SPL.

The task was a one-interval 4×8-alternative forced-choice with feedback (four colors: red, white, blue, and green and the numbers 1–8). Listeners were instructed to identify the color and number from the sentence with the callsign "Baron" and were informed that this target sentence would always be presented from the loudspeaker directly ahead. Since the target would always be directly ahead of them, they were told to keep their head facing forward, but they were not restrained. Responses were scored as correct only if

the listener identified both the color and number accurately. Listeners completed a short practice block of target identification in quiet at a comfortable listening level to familiarize them with the procedures and keypad.

2. Preliminary measurements

At the start of each session (for each room condition), a one-up, one-down adaptive procedure was used to estimate the 50% correct point on the psychometric function (Levitt, 1971) for speech identification in quiet (target only, no maskers). In these trials, the listener heard a single CRM sentence with the callsign "Baron" played from the loudspeaker at 0°. The initial step size was 4 dB and after four reversals the step size was decreased to 2 dB. Each adaptive track had a minimum of 30 trials and at least 9 reversals (typically many more than 9 were obtained). The threshold calculation was based on all reversals after the first three, discarding the fourth if necessary to make an even number. Two estimates of threshold in quiet were measured and averaged. If the threshold estimates differed by more than 5 dB an additional two estimates were collected and all four were averaged.

Following threshold measurements in quiet, percent correct speech identification was measured at a fixed target level (target only, no maskers). The groups with normal hearing were tested at 60 dB SPL. The groups with hearing loss were tested, if possible, at 30 dB sensation level (dB SL) with regard to their speech threshold in quiet. The upper limit for stimulus presentation was set at 84 dB SPL. With this in mind, the target level chosen for an individual listener with hearing loss was set high enough to be clearly understood, but low enough to allow the maskers to adapt over a sufficient range for estimating threshold. Some listeners with high quiet thresholds were tested at a lower SL to ensure that the upper limit for presentation was not exceeded by a high-level masker. In cases where the initial level chosen did not yield satisfactory identification in quiet (i.e., $\geq$87% correct), the level was increased slightly and fixed-level identification was repeated. For the BARE room conditions, four listeners were tested at 10 dB SL, two were tested at 15 dB SL, five were tested at 20 dB SL, and nine were tested at 30 dB SL. One of the listeners with hearing loss tested at 10 dB SL had slightly lower fixed-level identification (83%) than the other listeners; however, the level could not be increased further due to the equipment limits. Listeners were tested at the same sensation level in the PLEX room as in the BARE room with three exceptions. Two listeners with hearing loss were tested at slightly higher sensation levels in the PLEX room in order to achieve better fixed-level identification scores and one listener was tested at a lower sensation level in PLEX due to comfort. All listeners reported that the target levels used in the main experiment were at comfortable listening levels in both room conditions.

3. Speech thresholds with two competing talkers

The conditions comprising the main experiment were tested after the target level was chosen and confirmed to yield highly accurate identification. In these conditions, the listener heard three sources played concurrently (one target with two independent maskers) on every trial. The trials were blocked by masker location and presented in random order, where masker location changed after every two adaptive tracks. The maskers were either colocated with the target or symmetrically spatially separated (positioned at $\pm 90°$). The target level was fixed at the specified sensation level. The masker level was varied adaptively in 4 dB steps initially and then in 2 dB steps following the fourth reversal. The two masker talkers always had the same rms level. At the beginning of each adaptive track, the target was clearly audible above the maskers (the initial masker level was 20 dB below the target). As in the measurement of quiet threshold, each track had a minimum of 30 trials and at least 9 reversals and the threshold calculation was based on all reversals after the first three, discarding the fourth if necessary to make an even number. Threshold estimates were averaged over four tracks per condition.

III. RESULTS

A. Speech identification thresholds in quiet

The speech identification threshold measurements using the CRM stimuli that were made in both room conditions in quiet (without interfering talkers) were obtained for the purpose of setting the target level and are used in later analyses as an approximate estimate of the listener's amount of hearing loss (see Fig. 5). The older groups tended to have slightly higher thresholds in quiet than the younger groups. In both room conditions, threshold in quiet was on average 8 dB SPL higher in the ONH group than in the YNH group and 10 dB SPL higher in the OHL group than in the YHL group.

B. T/M at threshold with two competing talkers

To facilitate comparisons across conditions, masked thresholds were expressed as T/M in decibels. The T/M was calculated by subtracting the level of the individual maskers at threshold from the fixed target level. Group mean thresholds (and standard error of the mean) for target identification in the presence of two competing talkers for the two spatial separation conditions are shown in Fig. 2. The left panel contains results from the BARE room condition and the right panel shows results from the PLEX room condition. These data were analyzed using a repeated-measures analysis of variance (ANOVA) with two within-subjects factors (spatial separation and room condition) and two between-subjects factors (hearing status and age). Significant main effects were found for all four factors [spatial: $F(1,36)=236.9$, $p<0.001$; room: $F(1,36)=43.5$, $p<0.001$; hearing loss: $F(1,36)=36.9$, $p<0.001$; and age: $F(1,36)=14.5$, $p=0.001$] As shown in Fig. 2, the T/Ms at threshold were significantly lower (better) for (a) spatially separated versus colocated talkers, (b) the less reverberant room (BARE), (c) the listeners with normal hearing (open symbols), and (d) the younger listeners within each hearing status category (circles).

A comparison of data in the left and right panels of Fig. 2 illustrates the significant interactions between room condition and spatial separation [$F(1,36)=91.2$, $p<0.001$], room condition and hearing loss [$F(1,36)=9.0$, $p=0.005$],

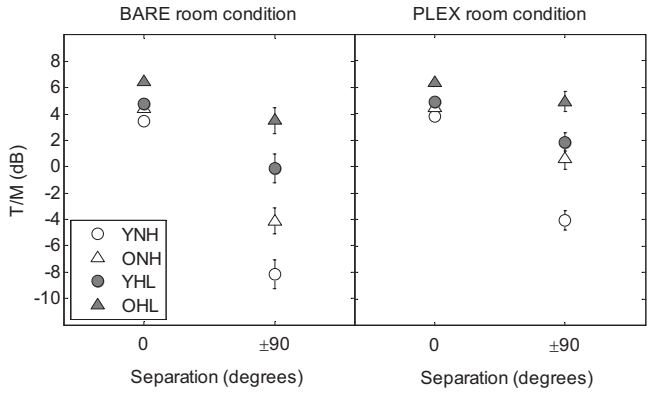


FIG. 2. Target-to-masker (T/M) ratio (expressed in decibels and computed relative to the level of individual maskers) at threshold in the two separation conditions (0° and ±90°) for each listener group. The error bars are ±1 standard error of the mean. The left panel contains results from the low-reverberant (BARE) room condition and the right panel shows results from the more reverberant (PLEX) room condition. Hearing status and age are denoted as shown in the symbol key.

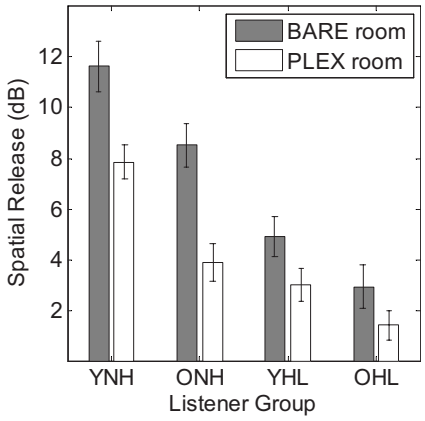


FIG. 3. Mean spatial release from masking in decibels (difference in T/M at threshold between colocated and separated conditions) for each listener group in both room conditions. The error bars are ±1 standard error of the mean.

and room condition, spatial separation, and hearing loss [$F(1,36)=16.7$, $p<0.001$]. Overall, the T/Ms at threshold were similar across room conditions when the talkers were colocated. Thresholds were at a target level that was slightly higher than that of the combined maskers (i.e., for a T/M at a threshold of 0 dB all talkers are equal in level whereas at +3 dB the target talker is 3 dB higher in level than each individual masker and approximately equal to the combined level of the two masker talkers). All eight mean thresholds (four groups by two rooms) occurred within a 2.9 dB range. For all groups, the average T/Ms at threshold when the talkers were spatially separated were lower than when they were colocated. The eight mean thresholds (four groups by two rooms) for the spatially separated talkers spanned an 11.7 dB range.

There were also significant interactions between spatial separation and hearing loss [$F(1,36)=47.0$, $p<0.001$] and spatial separation and age [$F(1,36)=13.9$, $p=0.001$], which are observable within each panel of Fig. 2. In each case, as pointed out above, the groups differed in performance very little for the colocated condition and much more for the spatially separated condition. All other interactions were nonsignificant. Despite the small range of performance in the colocated condition, within each room condition the groups were ordered YNH, ONH, YHL, and OHL going from the lowest to the highest T/M (average T/M at threshold in BARE: 3.5, 4.4, 4.8, and 6.4 dB and in PLEX: 3.8, 4.5, 4.9, and 6.3 dB, respectively). The same ordering of thresholds for the listener groups held for the much larger range of thresholds in the spatially separated condition (BARE: −8.2, −4.1, −0.1, and 3.5 dB; PLEX: −4.1, 0.6, 1.9, and 4.9 dB).

C. Benefit of spatial separation

A comparison of the results for the colocated condition with those for the spatially separated condition was summarized in terms of the amount of spatial release from masking, a measure of the benefit of spatial separation. This value was calculated for each listener by subtracting the T/M at threshold for the ±90° separation from the T/M at threshold for the condition where the talkers were colocated. Figure 3 displays listener group on the abscissa with the amount of spatial release from masking in decibels along the ordinate. The mean and standard errors are shown for each listener group for both room conditions. This figure illustrates that all listener groups showed a benefit of spatial separation between the target and competing talkers in both room conditions. However, the amount of benefit was dependent on the listener group and on the room condition. Again, the order was consistent and held for both rooms with the largest release observed for the YNH group followed by ONH, YHL, and OHL groups. Statistical analysis by repeated-measures ANOVA on spatial release from masking was completed with room condition as a within-subjects factor and with hearing status and age as between-subjects factors. The results revealed significant main effects of room reverberation [$F(1,36)=91.2$, $p<0.001$], hearing status [$F(1,36)=47.0$, $p<0.001$], and age [$F(1,36)=13.9$, $p=0.001$]. There was a significant interaction between room reverberation and hearing status [$F(1,36)=16.7$, $p<0.001$] but not between room and age [$F(1,36)=0.15$, $p=0.70$], or room, hearing status, and age [$F(1,36)=0.99$, $p=0.33$].

Individual listener data are shown in Fig. 4 comparing the amount of spatial release in the BARE room condition with the amount of spatial release in the PLEX room condition. The solid line is a least-squares fit to the data and the dashed line shows equivalence of the two measures. Most data points fall below the dashed line consistent with the group results, indicating that the amount of spatial release in the BARE room was greater than in the PLEX room for most individuals. There was also a significant correlation between the amount of spatial release in the BARE room and the amount of spatial release in the PLEX room ($r=0.86$, $p<0.001$). Individual differences in spatial release can be observed within all listener groups. Those listeners with less spatial release overall showed less difference between the two room conditions (compare distances from the dashed line), which illustrates the source of the statistical interaction between hearing status and room reverberation also seen in the mean results. Note also the overlap between listener groups in the range of spatial release from masking: While

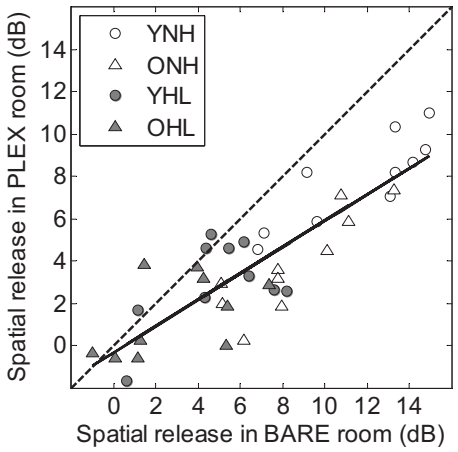


FIG. 4. Individual listener results for the amount of spatial release from masking in the PLEX room as a function of the amount of spatial release in the BARE room. The solid line is a least-squares fit to the data.

there was considerable overlap between younger and older listeners within each hearing status group there was much less overlap between listeners with and without hearing loss within each age group.

Figure 5 displays the relationship between spatial release from masking and speech identification thresholds in quiet (a measure related to the presence/magnitude of hearing loss). Results for the two room conditions are presented separately, with results for the BARE room condition in the left panel and results for the PLEX room condition in the right panel. The solid lines are least-squares fits to the data. Overall, the amount of spatial release from masking decreased with increasing quiet threshold. There was a strong and statistically significant negative correlation between quiet threshold and spatial release from masking in the BARE room condition ($r=-0.84$, $p<0.001$). The slope was such that spatial release from masking decreased by about 2 dB for every 10 dB increase in speech identification threshold in quiet. This relationship was reduced to a moderate but significant correlation in the PLEX room condition ($r=-0.66$, $p<0.001$). Because the listeners with the greatest amount of hearing loss were by necessity tested at a lower sensation level the influence of this factor should also be considered. Even if the six listeners who were tested at sensation levels below 20 dB in

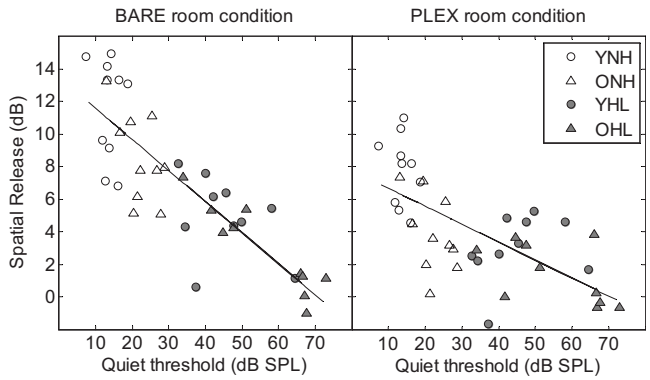


FIG. 5. Individual listener results showing the amount of spatial release from masking in decibels as a function of quiet threshold for identification of the CRM stimuli in each room condition. The solid lines are the least-squares fits to the data.

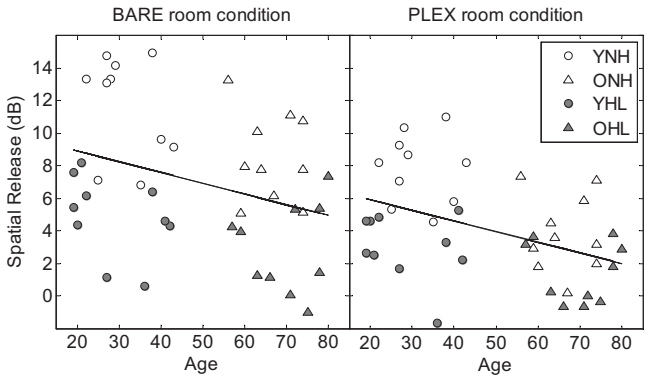


FIG. 6. Individual listener data for the amount of spatial release in decibels for each room condition as a function of age. The solid lines are the least-squares fits to the data.

the BARE room condition are excluded the correlation between quiet threshold and spatial release remains significant ($r=-0.72$, $p<0.001$ in BARE). For the PLEX case, the correlation excluding the five listeners who were tested at sensation levels of 15 dB or below is −0.60 ($p<0.001$). While it is prudent to keep in mind the potential confound of sensation level, these and other factors lead us to believe that it is not driving the effects seen here.[2]

The relationship between age and spatial release for individual listeners (coded by group) is shown in Fig. 6. The left panel displays the results in the BARE room condition and the right panel shows the results for the PLEX room condition. The solid lines are least-squares fits to the data. The negative correlations between age and spatial release were relatively weak but were significant in each room condition ($r=-0.32$, $p=0.044$ in BARE; $r=-0.43$, $p=0.005$ in PLEX). Because of the finding above, that just a 10 dB decrease in quiet threshold was associated with a 2 dB decrease in spatial release (for the BARE room condition), the difference in quiet thresholds across age groups needs to be taken into account. To help understand the relative contributions of these two variables, multiple regression analyses were performed with the amount of spatial release from masking for each room condition serving, in turn, as the predicted variable. Speech reception thresholds for the CRM stimuli in quiet and listener age were included as predictor variables. For the less reverberant room (BARE), 70% of the variance in the amount of spatial release from masking could be accounted for by variations in the speech identification threshold for the CRM stimuli in quiet ($p<0.001$). No significant increase in the variance accounted for by the model was achieved by the addition of age as a factor ($p=0.14$). For the more reverberant room (PLEX), the listener's speech reception threshold in quiet accounted for 44% of the variance ($p<0.001$) and including age as a factor increased the variance accounted for by 8% ($p=0.02$).

In summary, these results indicate that the listener's amount of hearing loss was the primary factor influencing the reduced benefit of spatial separation. This was less true as the amount of reverberation was increased. However, the effect of age, independent of hearing loss, appeared to show the opposite trend. The correlation and regression analyses

support a small, but statistically significant, effect of age on spatial release from masking in the more reverberant listening condition.

IV. DISCUSSION

The current study examined situations in which the listener must selectively attend to one talker in the presence of competing talkers. To do this, the listener must perceptually segregate the target speech stream and focus attention on it. There are several ways the listener could accomplish this task, including listening for the features of the talker's voice (e.g., fundamental frequency, prosody, and vocal tract length), using differences in level between talkers, following the content of the message, or using spatial cues. This study focused on the benefit of spatial information by holding all other factors constant or allowing them to vary in controlled but uninformative ways (e.g., randomizing talkers). In natural settings listeners typically monitor the entire auditory scene in order to switch attention to other sources when necessary, but for the current study they could focus exclusively on one spatial location. Because the target and masker sentences were spoken by same-sex talkers uttering highly similar and structured sentences, the task was subjectively quite difficult and made more so, we believe, by any degradation to the stimulus,[3] such as that caused by hearing loss and reverberation. The adverse effect of stimulus degradation was observed primarily in the spatially separated condition and that condition drove the magnitude of the difference underlying spatial release from masking. The similarity of the effect of hearing loss and reverberation on performance in the spatially separated condition suggests that other types of stimulus degradations (e.g., filtering, distorting, etc.) might likewise have an adverse effect on performance. However, the current results do not allow us to draw that conclusion and further study of the issue appears to be warranted.

A. Interpretation of colocated results

The threshold T/Ms for the target talker in the presence of two colocated masker talkers in the current study were comparable to the performance found by Brungart *et al.* (2001) for monaural segregation of three same-sex talkers using the same stimuli. The consistency of the thresholds across listeners and conditions (and studies, cf. Brungart *et al.*, 2001) was considered to be quite remarkable and suggests that the listeners have a reliable source segregation cue available. The most likely candidate, it seems to us, is the level of the target relative to the level of the maskers. According to this view, the presence of three simultaneous same-sex talkers obscures the talker-specific cues that the listener uses to follow the target voice from the callsign to the test words. The words that are reported are thus the more salient of those spoken. In this case, this occurs when the target level reaches a point where it is higher than the summed level of the two maskers (in our current metric the target level would equal the combined masker level at a T/M of +3 dB). Neither hearing loss nor reverberation appear to substantially affect this cue. This finding is in contrast to those of some other studies using only a single colocated masker talker in which listeners with hearing loss typically required a higher T/M than listeners with normal hearing for equivalent performance (cf. 12.8 dB higher in Duquesnoy, 1983; approximately 5 dB higher in Gelfand *et al.*, 1988; and 3.7 dB higher in Arbogast *et al.*, 2005). However, these studies primarily tested older adults with hearing loss and in the current study, that group tended to require a slightly higher T/M than the other listener groups (see Fig. 2). In the study by Kidd *et al.* (2005) using methods similar to those used in the present study, threshold T/Ms in the PLEX room condition were 7 dB higher than in the BARE room condition. The main difference between studies was that there was only one masker talker and the stimuli were processed into nonoverlapping narrow frequency bands. The processing of the speech into narrow mutually exclusive frequency bands substantially reduced the spectral overlap of target and masker (reducing energetic masking, pitch, and intonation as factors) and may have provided a strong timbre cue useful in source segregation. Because of those differences, perhaps, the threshold T/Ms for both colocated and separated conditions were considerably lower than here. Thus we think it is likely that the difference between the findings of the current study and previous studies may be attributed to a difference in the segregation cues available for distinguishing between two versus three speech sources. The possible influence of a level "boundary" where a change in strategy may improve performance has been suggested in other studies of informational masking using a variety of different paradigms (e.g., Richards and Neff, 2004; Arbogast *et al.*, 2005).

B. Interpretation of spatially separated results

The results of the current experiment help to clarify one reason why listeners with hearing loss report difficulty selectively attending to speech in a multitalker background: They appear not to be able to make use of spatial separation cues as well as the listeners with normal hearing. In fact, for some listeners with hearing loss, their performance when the talkers were spatially separated was no better than when the talkers were colocated. The normal-hearing listeners were able to reach negative T/Ms at threshold; that is, they could tolerate the interfering talkers when they were actually substantially higher in level than the attended target talker. However, the listeners with hearing loss needed a higher, typically positive, T/M to understand the target talker correctly 50% of the time. The practical implication of this finding is important since listeners with hearing loss would likely require higher T/Ms than listeners with normal hearing to communicate effectively in everyday listening situations. As a result, the listeners with hearing loss had significantly less spatial release from masking than the normal-hearing listeners since the colocated condition did not differ much across listener groups. There was a strong and significant negative correlation between the amount of hearing loss and spatial release from masking. The relationship was such that the higher the listener's quiet threshold (i.e., more hearing loss), the less they benefited from the spatial separation cues. In these conditions, we speculate that factors that degrade the

stimulus, in this case hearing loss and reverberation, affect the performance more for spatially separated sources than for colocated sources.

The magnitude of the benefit of spatial separation between sources in the current study was larger for the listeners with normal hearing than in previous studies using symmetrically placed noise maskers and the subsequent difference between listener groups was also larger. Bronkhorst and Plomp (1992) used speech modulated noise maskers in a variety of symmetric or asymmetric configurations that were recorded in an anechoic room and presented to the listeners under headphones either monaurally or binaurally. In the conditions that are most relevant to the current experiment (two noise maskers either colocated with the target at 0° or symmetrically separated at ±90° presented binaurally), spatially separating the maskers led to a 4.6 dB improvement in T/M for the normal-hearing listeners and a 2.7 dB improvement in T/M for the listeners with hearing loss. Ter-Horst *et al.* (1993) also found that listeners with hearing loss show less benefit than listeners with normal hearing from the separation of speech from two correlated noises shaped to match the long-term average speech spectrum. They tested symmetric configurations for both groups of listeners at ±18° and ±54° either horizontal or vertical separation. While the listeners with normal hearing had 4–6 dB spatial release for the horizontal symmetric separations, the listeners with hearing loss had around 1 dB spatial release on average.

Another study important for comparison to the current results is that of Peissig and Kollmeier (1997). Again, the trend in performance was similar to the current results but the size of the effect was smaller, which may be attributed to the emphasis on informational masking in the design of the current study. Peissig and Kollmeier (1997) compared the subjective intelligibility of two known target sentences in listeners with hearing loss to listeners with normal hearing in a virtual soundfield. In the condition using interfering speech tested in both listener groups, there were two fixed interfering sources (105° and 255°) and one variable interfering source (from 0° to 360°). For the five listeners with hearing loss that could be tested in this condition, they judged the target sentences to be 50% intelligible at signal-to-noise ratios that ranged from 0.6 to 3.8 dB lower when the variable interferer was spatially separated as compared to when it was colocated with the target. This benefit was less than that achieved by the normal-hearing listeners in the same condition who showed a maximum benefit of approximately 5 dB. Our results had the same pattern: Listeners with hearing loss showed a benefit of spatial separation between 0 and 8.2 dB while listeners with normal hearing obtained benefits up to a maximum of about 15 dB. There are, however, several important differences in the design of the two studies that could help to explain the differences in the size of the effects. First, the interfering sources in the current study were highly similar talkers uttering sentences that could plausibly be misidentified as targets. Second, there was also greater uncertainty in our procedure because the target talker and keywords varied from trial to trial, as did that of the other masker talkers. Finally, Peissig and Kollmeier (1997) used different-sex talkers (two male and two female) providing the opportunity for segregating sources unrelated to spatial separation.

C. Interpretation of the effect of hearing loss

Why should sensorineural hearing loss affect the ability of listeners to selectively attend to one talker in the presence of competing talkers? There are several possibilities that may not be mutually exclusive. A few of these are discussed here. It should be mentioned at the outset that achieving low thresholds for the task used in this study inherently depends on binaural information because no spatial release is obtained for monaural listening. As in the preceding study (Marrone *et al.*, 2008), our interpretation of the spatial release obtained in these conditions is that it is primarily due to source segregation and focus of attention at a point in space that are enhanced by binaural cues.

First, it is possible that hearing loss increases the amount of energetic masking that the listeners experience. The CRM materials and testing procedure are designed to emphasize informational masking (e.g., Brungart *et al.*, 2001; Arbogast *et al.*, 2002). Recently, Brungart *et al.* (2006) concluded that indeed informational masking normally dominates the masking observed in multitalker masking situations tested using the CRM. This conclusion was based on an analysis of the signal-to-noise ratio in spectrotemporal bins at various stimulus levels and the associated listener identification performance. However, the reduced frequency and temporal resolution associated with sensorineural hearing loss could produce greater energetic masking (spectrotemporal overlap) than observed in listeners with normal hearing. In support of that idea, Arbogast *et al.* (2005) found that there was greater energetic masking in listeners with hearing loss than in listeners with normal hearing. This difference was emphasized by the use of processing that rendered the target and masker speech into mutually exclusive sets of narrow bands. By comparing the masking produced by both narrow bands of noise and narrow bands of speech, they concluded that the reduced frequency resolution in the hearing-impaired listeners increased the ratio of energetic to informational masking. Because perceptual segregation cues are generally less effective for energetic masking than for informational masking (cf. Freyman *et al.*, 1999, 2001; Kidd *et al.*, 2008), less spatial release from masking would be an expected consequence of increased energetic masking.

A second possible explanation for the decreased benefit of spatial separation of sources observed in listeners with hearing loss is a reduction in the extent to which the sound sources are perceptually segregated when they are spatially separated. According to this argument, for the listeners with hearing loss, the images of the stimuli may be less "distinct" than for the normal-hearing listeners. This assumes that segregation is not an all or none phenomenon but instead varies in strength depending on the potency of the cues used to form different sound images. A related issue has to do with the usefulness of ILDs in sound segregation. Although there is no usable better ear, and the current results are not consistent with the hypothesis discussed above regarding reduced spatial release because of reduced better-ear listening, there

is the possibility that the moment-to-moment differences in level among the sources—indicating different source locations—aid in segregation. The observation that listeners with hearing loss usually lose more high frequency information where ILDs are most potent could mean that that binaural cue for sound segregation is compromised. Although the monaural control condition tested by Marrone *et al.* (2008) rules out monaural "glimpsing" of the target during epochs of high T/M, it is not conclusive with respect to the possible role of time-varying ILDs in providing segregation cues. Thus, resolution of this issue awaits further study.

There is evidence from other studies using different paradigms that listeners with hearing loss may have poorer source segregation abilities as compared to those with normal hearing. Using alternating tone sequences to test sequential stream segregation, Rose and Moore (1997) found that one-half of their listeners with bilateral hearing loss judged perceptual fission of the two streams when the frequency difference between them was much larger than that needed by their normal-hearing counterparts. In a similar paradigm, Mackersie *et al.* (2001) related the boundary for perceptual fusion of two tone sequences to the perception of simultaneous sentences. They found that listeners who judged that perceptual fusion occurred at smaller frequency differences tended to have better understanding of two simultaneous sentences. Grose and Hall (1996) measured sequential processing on a gap detection/discrimination task and a melody recognition task. While the listeners with hearing loss showed the same patterns of performance on the tasks as those with normal hearing, they performed more poorly on both tasks overall. This was interpreted as evidence that hearing loss affects the perceptual organization of sequential stimuli, possibly due to a need for greater frequency separation between auditory streams as compared to listeners with normal hearing. In a different paradigm, Kidd *et al.* (2002) found poorer perceptual segregation of individual components of complex sounds as a function of increasing amount of hearing loss. They measured detection of a pure tone sequence embedded in a multiple burst masker (either energetic or informational) as well as auditory filter measurements. The listeners with hearing loss showed a reduced ability to detect the target in a spectrotemporally complex and uncertain masker, a result which could not be explained by their auditory filter characteristics. This finding was instead attributed to a difference in analytic listening abilities, a conclusion that warrants consideration in the current study where analytic listening abilities are critical for hearing out one voice among several similar competing voices.

Finally, differences in other higher-level processes important to performing the task—such as attentional focus—between the groups could drive the observed differences in the benefit of spatial separation. If we assume that the different categories of listeners have the same pool of cognitive/attentional resources available, hearing loss may place an additional processing load on the observer leaving a smaller residual pool of resources available to solve the task. Consequently, if this conjecture were true, listeners with hearing loss would be predicted to perform equivalently to normal-hearing listeners when the selective listening task was relatively easy but would perform poorly when the task was difficult and required more processing resources. It seems reasonable to speculate that the listening environment becomes more complex and demanding when the number of distinct sources, the degree of listener uncertainty, and the distraction value of the sources increase. In the present study, all of these factors were, by design, chosen to be relatively high. A related but somewhat different explanation is that the listeners with hearing loss simply have difficulty inhibiting the processing of irrelevant stimuli. For example, Doherty and Lutfi (1999) found that listeners with sensorineural hearing loss had greater difficulty ignoring unwanted sounds as compared to normal-hearing listeners in a level discrimination task. Presumably, the greater the similarity of the target and maskers, the greater the demands on cognitive processing and the more difficult it is to ignore irrelevant stimuli. In the current experiment, the listeners with hearing loss informally reported that the task was quite effortful and often stated that this was because they had difficulty ignoring the masker talkers even when they were spatially separated.

D. Effect of reverberation

For all listener groups, performance in the spatially separated condition was also adversely influenced by reverberation. When both hearing loss and reverberation were present, performance was worse than with either alone. There was a relationship between the benefit of spatial separation observed in the more reverberant room and how much benefit was shown in the less reverberant room. Listeners demonstrating larger spatial release in the BARE room condition were those listeners with the larger spatial release in the PLEX room condition. These listeners also showed the greatest difference between the room conditions (i.e., greatest deviation from the diagonal line in Fig. 4). The present results differ somewhat from similar measures in normal-hearing listeners reported by Kidd *et al.* (2005). In that study, increased reverberation elevated speech identification thresholds in *both* colocated and spatially separated conditions about equally, so the large spatial release they found was preserved. As noted above, we believe that the reduction in spatial release found here is a consequence of the limited utility of normal talker segregation cues in the colocated condition leaving relative level as the primary, and robust, basis for selecting the target talker.

E. Effect of age

Although age was a significant factor in the overall results from the repeated-measures ANOVAs, this must be interpreted with caution due to the results of further analyses. The small difference in thresholds between age groups (cf. Fig. 1) is enough to account for part of the difference in spatial release as indicated by the lines representing the least-squares fits shown in Fig. 5. Furthermore, the correlations between spatial release and age were weak (cf. Fig. 6) and the regression analyses indicate no further contribution of age once quiet threshold is accounted for in the less reverberant room and only a small contribution (an additional 8% of the variance accounted for) in the more reverberant room.

Consequently, performance differences observed across listener groups cannot unambiguously be attributed to age effects independent of hearing status. This is consistent with the finding of Li *et al.* (2004), where younger and older adults with clinically normal hearing were tested in conditions in which the target and masker talkers were perceived to be originating from different locations (cf. Freyman *et al.*, 1999). Overall, the relative improvement due to perceived spatial separation was the same for younger and older adults, although the older adults needed a slightly higher T/M for equivalent performance. Dubno *et al.* (2008) also found that the older adults had generally higher thresholds but the same amount of benefit due to spatial separation for the speech in noise task using symmetric maskers. However, that benefit was smaller than their articulation-index based predictions. Interestingly, Humes *et al.* (2006) found that older adults performed significantly worse than younger adults on eight of nine divided attention tasks but on only three of nine selective attention tasks. This suggests that an age effect might be observable under the conditions of the current experiment if the task demands were more challenging; for example, if the keywords from two talkers must be reported on every trial. Recently, Helfer and Freyman (2008) compared speech-on speech masking in younger and older listeners under conditions in which differences in apparent spatial location of target and masker normally provide a significant enhancement in performance. Although several types of maskers were tested, of particular interest here was their finding that older listeners were less able to use spatial cues to improve speech recognition in general but especially for different-sex, rather than same-sex, talkers. This effect was particularly apparent in older listeners with hearing loss. Thus, a greater age effect might have occurred in this study had different-sex talkers been tested. The authors conclude that age-related changes in cognitive function may affect the ability to selectively attend to one talker among other spatially separated talkers, an effect that may be related to a general difficulty in ignoring irrelevant stimuli of high semantic content.

V. SUMMARY

Consistent with the report by Marrone *et al.* (2008), normal-hearing listeners demonstrated a large benefit of spatial separation when listening to multiple simultaneous talkers and trying to identify the speech of one of them. This effect was obtained without the availability of a better-ear advantage and was relatively robust with respect to increased reverberation. Bilateral sensorineural hearing loss decreased the benefit of spatial separation of sources. All listener groups had similar T/Ms at threshold when the three talkers were colocated. However, when the talkers were spatially separated, listeners with hearing loss required much higher T/Ms at threshold than normal-hearing listeners, particularly in the reverberant environment. For some listeners with hearing loss, performance was as poor in the spatially separated condition as in the colocated condition. The effect of age separate from hearing loss, however, was inconclusive. Increased energetic masking and decreased ability to segregate and focus attention in space were considered possible explanations for the findings. Overall, these results help to explain why listeners with hearing loss often report great difficulty understanding a target talker in real rooms when there are other talkers in the background.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. DC004545, DC00100, and DC004663 from the National Institute on Deafness and Other Communication Disorders (NIDCD). N.M. was supported by a Ruth L. Kirschstein National Research Service Award for predoctoral fellows from NIDCD (F31 DC008440). This work partly fulfills the requirements for the Doctor of Philosophy degree at Boston University by N.M. We are grateful to Deborah Corliss and Janine Francis for their assistance in the Sound Field Laboratory. Thanks also to Virginia Best, Barbara Shinn-Cunningham, Melanie Matthies, Steven Colburn, Richard Freyman, and three anonymous reviewers for helpful comments on earlier versions of this manuscript.

[1]The measurements as described in Kidd *et al.*, 2005 pertain directly here as they involve the exact same surface treatments, loudspeakers and loudspeaker positions, listener position, etc., as in the current experiment. The purpose was to compare across our own experimental room conditions; therefore, generalization to other rooms and different measurement procedures (e.g., omnidirectional source) should be made with caution.

[2]To obtain more insight into the possible effects of sensation level on spatial release from masking several listeners were retested at different sensation levels in the BARE room condition using methods identical to the main experiment. When four listeners from the YNH group were retested at 20 dB SL (equal to or less than that used with most listeners in the hearing loss groups), they continued to show large spatial release from masking. In addition, when two listeners from the YHL group were tested at higher sensation levels their spatial release from masking did not improve. However, when the four YNH listeners and four additional YHL listeners were tested at 10 dB SL, the amount of spatial release was reduced but not eliminated on an individual listener basis within both groups. Although these individuals may not be representative of the larger groups we believe that this in addition to the fact that the correlations remain significant between quiet threshold and spatial release when the low SL listeners are excluded is convincing evidence that the different SLs used are not the reason for the effect.

[3]One effect of reverberation is the disruption of the binaural cues of ILD and ITD, which we include in our characterization of stimulus degradation. As discussed in the text, cochlear hearing loss also may be thought of as causing a degradation of the stimulus.

American National Standards Institute. (**2004**). American national standard specification for audiometers. ANSI 3.6-2004. New York, NY: Acoustical Society of America.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2005**)."The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Best, V., Gallun, F. J., Ihlefeld, A., and Shinn-Cunningham, B. G. (**2006**). "The influence of spatial separation on divided listening," J. Acoust. Soc. Am. **120**, 1506–1516.

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.

Bronkhorst, A. W., and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1374–1383.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Bronkhorst, A. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acust. Acta Acust. **86**, 117–128.
Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.
Brungart, D. S., and Simpson, B. D. (**2002**). "Within-ear and across-ear interference in a cocktail-party listening task," J. Acoust. Soc. Am. **112**, 2985–2995.
Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.
Colburn, H. S., Shinn-Cunningham, B. A., Kidd, G., Jr., and Durlach, N. I. (**2006**). "The perceptual consequences of binaural hearing," Int. J. Audiol. **45**, 34–44.
Doherty, K. A., and Lutfi, R. A. (**1999**). "Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **105**, 1831–1840.
Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (**2002**). "Spectral contributions to the benefit from spatial separation of speech and noise," J. Speech Lang. Hear. Res. **45**, 1297–1310.
Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (**2008**). "Binaural advantage for younger and older adults with normal hearing," J. Speech Lang. Hear. Res. **51**, 539–556.
Duquesnoy, A. J. (**1983**). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739–743.
Ebata, M. (**2003**). "Spatial unmasking and attention related to the cocktail party problem," Acoust. Sci. & Tech. **24**, 208–219.
Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.
Gatehouse, S., and Noble, W. (**2004**). "The speech, spatial, and qualities of hearing scale (SSQ)," Int. J. Audiol. **43**, 85–99.
Gelfand, S. A., Ross, L., and Miller, S. (**1988**). "Sentence reception in noise from one versus two sources: Effects of aging and hearing loss," J. Acoust. Soc. Am. **83**, 248–256.
Grose, J. H., and Hall, J. W. (**1996**) "Perceptual organization of sequential stimuli in listeners with cochlear hearing loss," J. Speech Hear. Res. **39**, 1149–1158.
Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in complex environments," J. Acoust. Soc. Am. **105**, 3436–3448.
Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.
Helfer, K. S. (**1992**). "Aging and the binaural advantage in reverberation and noise," J. Speech Hear. Res. **35**, 1394–1401.
Helfer, K. S., and Freyman, R. L. (**2008**). "Aging and speech-on-speech masking," Ear Hear. **29**, 87–98.
Humes, L. E., Lee, J. H., and Coughlin, M. P. (**2006**). "Auditory measures of selective and divided attention in young and older adults using single-talker competition," J. Acoust. Soc. Am. **120**, 2926–2937.
Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**) "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.
Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Walsh, M. (**2002**). "Informational masking in listeners with sensorineural hearing loss," J. Assoc. Res. Otolaryngol. **3**, 107–119.
Kidd, G., Jr., Mason, C. R., Brughera, A., and Hartmann, W. M. (**2005**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acust. Acta Acust. **91**, 526–536.
Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. (**2008**). "Informational masking" in *Auditory Perception of Sound Sources*, edited by W. A. Yost (Springer, New York), pp. 143–189.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Li, L., Daneman, M., Qi, J. G., and Schneider, B. A. (**2004**). "Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults?," J. Exp. Psychol. Hum. Percept. Perform. **30**, 1077–1091.
Mackersie, C. L., Prida, T. L., and Stiles, D. (**2001**). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **44**, 19–28.
Marrone, N., Mason, C. R., and Kidd, G., Jr. (**2008**). "Tuning in the spatial dimension: Evidence from a masked speech identification task," J. Acoust. Soc. Am. **124**, 1146–1158.
Noble, W., and Perrett, S. (**2002**). "Hearing speech against spatially separate competing speech versus competing noise," Percept. Psychophys. **64**, 1325–1326.
Peissig, J., and Kollmeier, B. (**1997**)."Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670.
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.
Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.
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.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, and Larson, E. (**2005**). "Bottom-up and top-down influences on spatial unmasking," Acust. Acta Acust. **91**, 967–979.
Ter-Horst, K., Byrne, D., and Noble, W. (**1993**). "Ability of hearing-impaired listeners to benefit from separation of speech in noise," Aust. J. Audiol. **15**, 71–84.
Yost, W. A. (**1997**). "The cocktail party problem: Forty years later" in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. A. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 329–348.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility" in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. A. Studebaker and I. Hockberg (Allyn and Bacon, Needham Heights, MA), pp. 255–276.

Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation[a)]

Etienne Gaudrain[b)] and Nicolas Grimault[c)]
Neurosciences Sensorielles, Comportement, Cognition, CNRS UMR 5020, Université Lyon 1, 50 Avenue Tony Garnier, 69366 Lyon Cedex 07, France

Eric W. Healy[d)]
Speech Psychoacoustics Laboratory, Department of Communication Sciences and Disorders, University of South Carolina, Columbia, South Carolina 29208

Jean-Christophe Béra
Inserm U556, 151, cours Albert Thomas, 69424 Lyon Cedex 03, France

(Received 17 December 2007; revised 21 August 2008; accepted 22 August 2008)

Cochlear-implant (CI) users often have difficulties perceiving speech in noisy environments. Although this problem likely involves auditory scene analysis, few studies have examined sequential segregation in CI listening situations. The present study aims to assess the possible role of fundamental frequency (F_0) cues for the segregation of vowel sequences, using a noise-excited envelope vocoder that simulates certain aspects of CI stimulation. Obligatory streaming was evaluated using an order-naming task in two experiments involving normal-hearing subjects. In the first experiment, it was found that streaming did not occur based on F_0 cues when natural-duration vowels were processed to reduce spectral cues using the vocoder. In the second experiment, shorter duration vowels were used to enhance streaming. Under these conditions, F_0-related streaming appeared even when vowels were processed to reduce spectral cues. However, the observed segregation could not be convincingly attributed to temporal periodicity cues. A subsequent analysis of the stimuli revealed that an F_0-related spectral cue could have elicited the observed segregation. Thus, streaming under conditions of severely reduced spectral cues, such as those associated with CIs, may potentially occur as a result of this particular cue. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2988289]

PACS number(s): 43.66.Mk, 43.66.Sr, 43.71.Es, 43.71.Ky [JCM] Pages: 3076–3087

I. INTRODUCTION

The mechanisms involved in auditory stream segregation have been thoroughly investigated in normal-hearing (NH) listeners (e.g., Bregman and Campbell, 1971; van Noorden, 1975; Bregman, 1990). These studies led to the *peripheral channeling theory* (Hartmann and Johnson, 1991), which states that two stimuli need to excite different peripheral neural populations to produce auditory streaming. This theory and its implementations (Beauvois and Meddis, 1996; McCabe and Denham, 1997) assume that the main cues for streaming are spectral, suggesting that frequency selectivity is critical. Moore and Gockel (2002), in a review of studies involving sequential stream segregation, further concluded that *any* sufficiently salient perceptual difference may lead to stream segregation, regardless of whether or not it involves peripheral channeling (see also Elhilali and Shamma, 2007). Frequency selectivity can also affect the perceptual salience of cues, and difference limen (DL) measurements can be used to evaluate the salience of stimuli along a given perceptual dimension. Rose and Moore (2005) tested this hypothesis and found that the fission boundary (cf. van Noorden, 1975) was indeed proportional to the frequency DL for pure tones between 250 and 2000 Hz. However, it can be difficult to clearly define the salience of complex sounds composed of many interacting features. Moreover, this difficulty can be compounded when the signal is degraded by the hearing system, such as in hearing-impaired (HI) listeners or in cochlear-implant (CI) users. The current study aims to clarify the role of fundamental frequency in the perceptual segregation of vowel sequences having spectral cues reduced through the use of an acoustic vocoder model of a CI (cf. Dudley, 1939; Shannon *et al.*, 1995).

Experiments involving NH listeners have shed light on the mechanisms underlying pitch-based streaming and on the influence of reduced frequency resolution. Streaming based on fundamental frequency (F_0) is reduced when the resolvability of harmonic components of complex tones is reduced, but it is still possible to some extent even when harmonics are totally unresolved (Vliegen and Oxenham, 1999; Vliegen *et al.*, 1999; Grimault *et al.*, 2000). Gaudrain *et al.* (2007)

[a)]Portions of this work were presented in "Segregation of vowel sequences having spectral cues reduced using a noise-band vocoder." Poster presented at the 151st ASA meeting in Providence, RI, June 2006.
[b)]Present address: Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom.
[c)]Electronic mail: ngrimault@olfac.univ-lyon1.fr
[d)]Present address: Department of Speech and Hearing Science, The Ohio State University, Columbus, 43210.

 © 2008 Acoustical Society of America

found that F_0-based streaming of vowel sequences was reduced when frequency resolution was reduced by simulated broad auditory filters (Baer and Moore, 1993). Roberts *et al.* (2002) showed that differences solely in temporal cues (obtained by manipulating the phase relationship between components) can elicit streaming. Finally, Grimault *et al.* (2002) observed streaming based on the modulation rate of sinusoidally amplitude-modulated noises, i.e., without any spectral cues to pitch. Despite the fact that the pitch elicited by the modulated noise was relatively weak, these authors observed streaming similar to that obtained with unresolved complex tones. Thus, streaming is reduced when spectral cues are reduced, but it is apparently possible to some extent when spectral cues are removed and only temporal cues remain.

These results have substantial implications for individuals with sensorineural hearing impairment and those fitted with a CI. It is well known that these individuals have reduced access to spectral cues (cf. Moore, 1998). Fundamental frequency DLs are approximately 2.5 times greater than normal in HI listeners (Moore and Peters, 1992) and 7.7 times greater in CI users (Rogers *et al.*, 2006). These results suggest that pitch differences are far less salient for these listeners, and that pitch-based streaming might be impaired. Indeed a few studies argue that reduced frequency selectivity is responsible for the relatively poor performance of CI users in the perception of concurrent voices (Qin and Oxenham, 2005; Stickney *et al.*, 2004, 2007). However, psychoacoustic measures have indicated that temporal resolution is generally intact in the HI ear (for review, see Moore, 1998; Healy and Bacon, 2002). CI users are also sensitive to temporal rate pitch, up to a limit of approximately 300 Hz (Shannon, 1983; Tong and Clark, 1985; Townshend *et al.*, 1987). Although their DLs for rate discrimination are larger than in NH (Zeng, 2002) CI listeners can use this cue to discriminate vowel F_0's (Geurts and Wouters, 2001). Although these results indicate that the cues for streaming may be available, their use by these individuals is not well understood.

Auditory streaming has been examined to a limited extent in HI listeners, with mixed results. Grose and Hall (1996) found that listeners with cochlear hearing loss generally required a greater frequency separation for segregation of pure tones. However, Rose and Moore (1997) reported no systematic difference between ears of unilaterally impaired listeners in this task. The correlation between auditory filter width and pure-tone streaming was also found to be not significant (Mackersie *et al.*, 2001). Grimault *et al.* (2001) found that streaming was hindered for HI listeners relative to NH, but only in conditions where components of complex tones were resolved for NH and unresolved for HI listeners. Finally, Stainsby *et al.* (2004) examined streaming based on phase relationship differences and found results for elderly HI listeners that were similar to those observed in NH listeners.

A few studies have also attempted to examine streaming in CI users (Hong and Turner, 2006; Chatterjee *et al.*, 2006; Cooper and Roberts, 2007). For these users, different kinds of temporal cues can be related to pitch. Moore and Carlyon (2005) argued that the temporal fine structure of resolved harmonics was the most accurate pitch mechanism. However, when harmonics are unresolved, they interact in auditory filters and can encode pitch by amplitude modulation (AM) rate (i.e., by the temporal envelope periodicity). Because of the way the spectrum is partitioned in the CI processor, harmonics of lower pitched human voices ($F_0 \sim 100$ Hz) almost always interact in the first channel of the CI processor. Thus, the availability of individual resolved harmonics is extremely limited. In contrast, the temporal envelope is roughly preserved in each band, so pitch may be coded by temporal periodicity cues. In this paper, the term "temporal-pitch cues" will then refer to temporal periodicity (in the range 100–400 Hz); in contrast to "spectral-pitch cues," which will refer to the pitch that is evoked by resolved harmonics (i.e., issued from the tonotopic analysis of the cochlea, and if relevant, from some analysis of temporal fine structure). Because amplitude-modulated broadband noises can produce some impression of pitch (Burns and Viemeister, 1976, 1981) and can induce streaming (Grimault *et al.*, 2002), it might be possible for these temporal cues to induce streaming in CI users.

Hong and Turner (2006) used the rhythm task described in Roberts *et al.* (2002) to obtain an objective measure of obligatory streaming in NH and CI users. They found that half of the 16–22 electrode CI users performed as poorly as the NH listeners (suggesting streaming), whereas the other half performed better than normal (suggesting less stream segregation). The authors showed that this variability correlated moderately but significantly with the ability to perceive speech in noise. Chatterjee *et al.* (2006) used pulse trains in *ABA* patterns and a subjective evaluation of whether subjects fitted with the 22-channel nucleus CI heard one or two streams. These authors observed response patterns that could be explained by streaming for both differences in spatial location (presentation electrode) and AM rate (in a single subject). However, they did not observe the characteristic buildup of streaming over time (Bregman, 1978) for simple pulsatile stimuli that differed in location. This observation raises the possibility that the task involved discrimination rather than streaming. On the contrary, they did observe some buildup for the signals that differed in AM rate, which suggests that AM rate based streaming was indeed observed. Cooper and Roberts (2007) also employed pulsatile stimuli that differed in electrode location. They obtained subjective reports involving the presence of two streams, but a second experiment revealed that the results may have been attributable to pitch (or brightness) discrimination. Other studies have targeted temporal cues more specifically, but have examined simultaneous segregation by CI users. Using a CI simulation based on filtered harmonic complexes, Deeks and Carlyon (2004) found only modest improvements in concurrent sentence recognition when the target and masker were presented at different pulse rates. Also, Carlyon *et al.* (2007) found that a difference in rate pitch did not enhance simultaneous segregation of pulse trains in CI users. Altogether, these studies provide only modest evidence that segregation or streaming can occur in CI recipients on the basis of either place pitch (i.e., electrode number) or temporal pitch.

These previous results together suggest (1) that F_0-based streaming is affected by frequency selectivity, but (2) that

streaming can be also induced by temporal-pitch cues. It is also clear that (3) frequency selectivity is reduced in HI and CI listeners, but that (4) temporal-pitch cues are preserved to some extent in these listeners. The question then becomes to what extent these cues can be utilized to elicit streaming.

Although streaming is often assumed to be a primitive mechanism, some correlation between streaming and higher level processing, such as concurrent speech segregation, has been reported (Mackersie *et al.*, 2001). However, the relation between streaming with pure or complex tones and speech segregation remains difficult to assess. In speech, pitch cues signaling that different talkers are present are mixed with other cues that may not be relevant for concurrent talker segregation. Listeners may then not benefit from these cues in ecological situations. Only a few studies have reported streaming with speech materials (Dorman *et al.*, 1975; Nooteboom *et al.*, 1978; Tsuzaki *et al.*, 2007; Gaudrain *et al.*, 2007), and only the last one examined the effect of impaired frequency selectivity.

The current study follows that of Gaudrain *et al.* (2007). Whereas streaming under conditions of broad tuning in accord with sensorineural hearing impairment was examined in that study, streaming under conditions similar to CI stimulation was assessed in the current study. Specifically, the role of reduced spectral-pitch cues in streaming of speech stimuli was assessed in a first experiment, and the possible role of temporal-pitch cues was investigated in a second experiment.

In this study, noise-band vocoder models of CIs were employed to control the spectral and temporal cues available to listeners. The use of NH listeners exposed to cues reduced in controlled manner allowed the elimination of complications and confounds associated with the clinical population. In addition, an objective paradigm—the order task—was used to assess streaming. In this task, the listener is presented a repeating sequence of vowels having alternating F_0 and asked to report the order of appearance of the constituent vowels (Dorman *et al.*, 1975). If the sequence splits into streams corresponding to the two F_0's, the loss of temporal coherence across streams hinders the ability to identify the order of items within the sequence. Although the order of items within individual steams is available to the listener, the order of items across streams is not. Thus, this task requires that the subject *resists* segregation to perform well. As a result, the task is used to assess "obligatory" streaming—that which cannot be suppressed by the listener (see Gaudrain *et al.*, 2007).

This type of streaming does not produce a substantial cognitive load, and it is less dependent on attention and subject strategy than the subjective evaluation of one versus two streams. In addition, this approach is appropriate for examining segregation in the presence of reduced cues—because performance tends to improve (less streaming) with degraded stimuli, performance cannot be attributable to the degradation of the individual items. The reduction in spectral resolution associated with the vocoder is expected to reduce the amount of streaming. Consequently, performance in the order task should improve in the CI simulation, relative to the intact stimuli. However, if temporal-pitch cues encoded by the CI simulation are sufficient to elicit obligatory streaming, overall scores should remain low and an effect of F_0 separation should be observed.

II. EXPERIMENT 1

A. Materials and method

1. Subjects

Six subjects aged 22–30 years (mean 26.2) participated. All were native speakers of French and had pure-tone audiometric thresholds below 20 dB hearing level (HL) at octave frequencies between 250 and 4000 Hz (American National Standards Institute, 2004). All were paid an hourly wage for participation. These subjects participated in one of the experiments of Gaudrain *et al.* (2007) and were therefore familiar with the paradigm.

2. Stimuli

Individual vowels were first recorded and processed, then arranged into sequences. The six French vowels /a e i ɔ y u/ were recorded (24 bits, 48 kHz) using a Røde NT1 microphone, a Behringer Ultragain preamplifier, a Digigram VxPocket 440 soundcard, and a PC. The speaker was instructed to pronounce all six vowels at the same pitch and to reduce prosodic variations. The F_0 and duration of each vowel were then manipulated using STRAIGHT (Kawahara *et al.*, 1999). Duration was set to 167 ms to produce a speech rate of 6.0 vowel/s. This value is close to that measured by Patel *et al.* (2006) for syllable rates in British English (5.8 syllables/s) and in French (6.1 syllables/s). Additional versions of each vowel were then prepared in which the average F_0's were 100, 110, 132, 162, and 240 Hz. Fundamental frequency variations related to intonation were constrained to be within 0.7 semitones (4%) of the average. This value was chosen to allow F_0 variations within each vowel, but to avoid overlap across the F_0 conditions. Formant positions were held constant across F_0 conditions.

Each vowel was subjected to two conditions of reduced spectral resolution. In Q_{20} the vowels were subjected to a 20-band noise vocoder, and in Q_{12} they were subjected to a 12-band noise vocoder. The Q_{12} condition was intended to be closer to actual CI characteristics, while the Q_{20} condition was intended to be an intermediate condition with more spectral detail. Q_∞ refers to the intact vowels. The implementation of the noise-band vocoder followed Dorman *et al.* (1997). The stimulus was first divided into frequency bands using eighth order Butterworth filters. The cutoff frequencies of these bands were the approximately logarithmic values used by Dorman *et al.* (1998) and are listed in Table I. The envelope of each band was extracted using half-wave rectification and eighth order Butterworth lowpass filtering with cutoff frequency of 400 Hz. This lowpass value ensured that temporal-pitch cues associated with voicing were preserved. The resulting envelopes were used to modulate white noises using sample point-by-point multiplication, which were then filtered to restrict them to the spectral band of origin. The 12 or 20 bands comprising a condition were then mixed to construct the vocoder. A 10 ms cosine rise/fall gate was finally applied to each vowel in each condition.

TABLE I. Cutoff frequencies (Hz) of the 12- and 20-channel vocoders, from Dorman *et al.* (1998).

Channel	Q_{20}	Q_{12}
1	166	212
2	198	336
3	362	570
4	414	754
5	600	1056
6	676	1324
7	888	1718
8	992	2098
9	1236	2620
10	1376	3150
11	1658	3848
12	1840	4582
13	2168	
14	2400	
15	2786	
16	3078	
17	3534	
18	3900	
19	4438	
20	4894	
	5532	

The vowels were then concatenated to form sequences. Figure 1 describes the arrangement of vowels into sequences and the construction of the various conditions. Each sequence contained one presentation of each vowel. Sequences containing all possible arrangements of the six repeating vowels ($[n-1]!=120$) were first generated, then the 60 arrangements having the smallest differences in formant structure were selected for inclusion [Fig. 1(A)]. The selection of arrangements having the smallest *perceptual formant differences*[1] was performed to reduce the influence of streaming based on differences in formant structure between successive vowels in a sequence (Gaudrain *et al.*, 2007). These 60 arrangements were then divided into five groups of 12 arrangements each, such that the average perceptual distance of each group was approximately equal [Fig. 1(B)].

The F_0 of the vowels in a sequence alternated between two values $F_{0(1)}$ and $F_{0(2)}$. In condition LowRef, the value of $F_{0(1)}$ was 100 Hz and $F_{0(2)}$ was one of the five F_0 values (100, 110, 132, 162, and 240). In condition HiRef, $F_{0(1)}$ was 240 Hz and $F_{0(2)}$ was one of the five F_0 values. Thus, there were five F_0 differences. Each group of 12 arrangements was then assigned to one of the five F_0 differences. The appearance of the same 60 arrangements in both the LowRef and HiRef conditions yielded 120 sequences [Fig. 1(C)]. These 120 sequences appeared in both a Slow and a Fast condition, yielding 240 sequences [Fig. 1(D)]. Finally, each of these 240 sequences appeared in each of the three Q conditions, yielding a total of 720 sequences [Fig. 1(E)].

In the Slow condition, the presentation rate was 1.2 vowel/s, and in the Fast condition, it was 6 vowel/s. Slow sequences were used to check vowel identification performance, and Fast sequences were used to examine streaming.

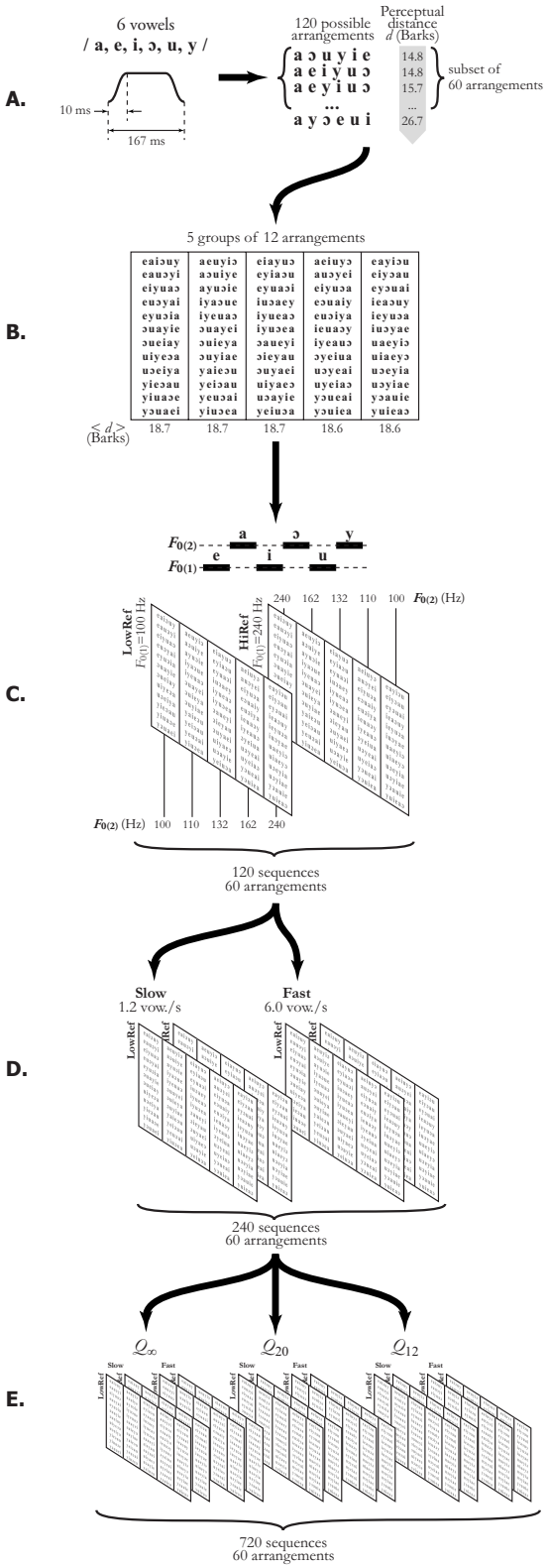


FIG. 1. The arrangement of conditions: (A) Individual vowels were recorded and modified using STRAIGHT. They were arranged into sequences (120 possible orders), and the 60 arrangements having the lowest perceptual distances d (in barks) were selected (see text for details). (B) These 60 arrangements were divided into five groups with similar average perceptual distance (12 in each group). (C) Each group was assigned to a fundamental frequency difference in both LowRef and HiRef conditions (yielding 120 sequences). (D) These 120 sequences appeared in both Slow and Fast conditions (yielding 240 sequences). (E) Finally, these 240 sequences appeared in each Q condition (yielding 720 sequences). These 720 sequences were presented across six presentation blocks, such that each condition was equally represented within each block.

To create the Slow sequences, silence was added between the vowels so that vowel duration remained constant across rate conditions. The Slow sequences were repeated four times and the Fast sequences were repeated 20 times, for overall stimulus durations of 20 s.

Stimuli were generated at 16 bits and 44.1 kHz using MATLAB. They were presented using the Digigram VxPocket 440 soundcard, and Sennheiser HD250 Linear II headphones diotically at 85 dB sound pressure level, as measured in an artificial ear (Larson Davis AEC101 and 824; American National Standards Institute, 1995).

3. Procedure

a. Training and selection. Two training tasks preceded testing. The first involved simple identification of single vowels. Subjects heard blocks containing each vowel at each F_0 twice. They responded using a mouse and a computer screen, and visual feedback was provided after each response. This test was repeated, separately for each Q condition, until a score of 98% (59/60) was obtained. On average, proficiency was reached after one block for the Q_∞ vowels, 1.3 blocks for the Q_{20} vowels, and 2.8 blocks for the Q_{12} vowels.

The second training task involved vowel identification using the Slow sequences. In each block, 60 sequences were presented representing all 30 conditions (5 F_0's $\times 2$ LowRef/HiRef $\times 3$ Q conditions). The procedure was the same as the test procedure, except that visual feedback was provided. The subject was presented with a repeating sequence. After an initial 5 s period, during which streaming was allowed to stabilize, the subject was asked to report the order of appearance of the constituent vowels. They were allowed to start with any vowel. The response was entered using a computer graphic interface and a mouse. The next sequence was presented after the subject confirmed their response or after a maximum of 20 s. Visual feedback was then provided. To proceed to the test, subjects were required to obtain a score, averaged over two consecutive blocks, greater than 95% in each Q condition. On average, 6.3 blocks were necessary to reach the proficiency criterion. Although intended to be a selection criterion, no subject was eliminated at this step.

b. Streaming test. The procedure was the same as that in the second training task, except that no feedback was provided. The 720 sequences were distributed among six presentation blocks, such that each condition was represented equally in each block. The average duration of one block was approximately 28 min. Experiment 1 required subjects to participate in four 2 h sessions, during which frequent breaks were provided. The experimental procedure was formally approved by a local ethics committee (CCPPRB Léon Bérard).

B. Results

For each condition, the score is the percentage of responses in which the six vowels comprising a sequence were reported in the correct order. Mean scores across subjects are plotted as a function of $F_{0(2)}$ in Fig. 2. Chance performance is 0.8%. As in Gaudrain *et al.*, 2007, high scores can be interpreted as a tendency toward integration across $F_{0(1)}$ and $F_{0(2)}$ items and a resistance to obligatory streaming. Separate analyses were conducted on the LowRef and HiRef conditions because the F_0 differences were not the same in the two conditions. The results in the Slow condition (1.2 vowel/s) showed that identification was near perfect in all conditions except one (HiRef, Q_{12}, $F_{0(2)}=162$ Hz). An analysis of errors in this condition showed confusions between /y/ and /e/ in 8/9 false responses. These two vowels therefore seem difficult to discriminate at this particular combination of F_0's and vocoder channel divisions. All subsequent analyses were carried out on the data collected in the Fast conditions.

A two-way analysis of variance (ANOVA) on the Fast/LowRef data using Q condition and $F_{0(2)}$ as repeated parameters indicated that the effects of Q condition [$F(2,10)=4.14$, $p<0.05$] and $F_{0(2)}$ [$F(4,20)=12.21$, $p<0.001$] were significant, and interacted significantly [$F(8,40)=3.93$, $p<0.01$]. Separate one-way ANOVAs on each Q condition using $F_{0(2)}$ as a repeated factor showed a significant effect of $F_{0(2)}$ in the Q_∞ condition [$F(4,20)=9.28$, $p<0.001$], but not in the Q_{12} [$F(4,20)=0.65$, $p=0.63$] or Q_{20} conditions [$F(4,20)=0.21$, $p=0.93$].

A two-way ANOVA on the Fast/HiRef data using Q condition and $F_{0(2)}$ as repeated parameters indicated that Q condition [$F(2,10)=9.02$, $p<0.01$] and $F_{0(2)}$ [$F(4,20)=14.30$, $p<0.001$] were significant, and interacted significantly [$F(8,40)=6.74$, $p<0.001$]. Separate one-way ANOVAs on each Q condition using $F_{0(2)}$ as a repeated factor showed significant effects of $F_{0(2)}$ in the Q_∞ condition [$F(4,20)=10.15$, $p<0.001$] and in the Q_{12} condition [$F(4,20)=14.96$, $p<0.001$], but not in the Q_{20} condition [$F(4,20)=1.50$, $p=0.24$]. A *post hoc* analysis using pairwise t-tests showed that the effect of $F_{0(2)}$ in the Q_{12} condition was due solely to the point $F_{0(2)}=162$ Hz. As previously stated, the confusions in this particular condition suggest difficulty with this particular set of parameters. When the HiRef condition was analyzed with this condition excluded, the pattern of significance was identical to that observed in the LowRef conditions: a significant effect of $F_{0(2)}$ in the Q_∞ condition [$F(3,15)=12.24$, $p<0.001$], but not in the Q_{12} [$F(3,15)=0.74$, $p=0.54$] or Q_{20} conditions [$F(3,15)=1.13$, $p=0.37$].

C. Discussion

The results in the natural speech condition (Q_∞) are consistent with those observed by Gaudrain *et al.* (2007) in their first experiment. The greater the F_0 difference, the lower the scores, signifying greater streaming. Streaming based on F_0 difference is considered to be *obligatory* here because the task employed required that streaming be suppressed in order to perform accurately. Although the pattern of results in the current experiment is similar to that obtained by Gaudrain *et al.* (2007), the baseline level of performance differs. Scores in the Q_∞ condition at matched F_0 were over 80% here and approximately 50% in experiment 1 of Gaudrain *et al.* (2007). One possible reason is that participants in the current experiment were well trained. In addition, there were subtle differences in the stimuli used in the two studies.

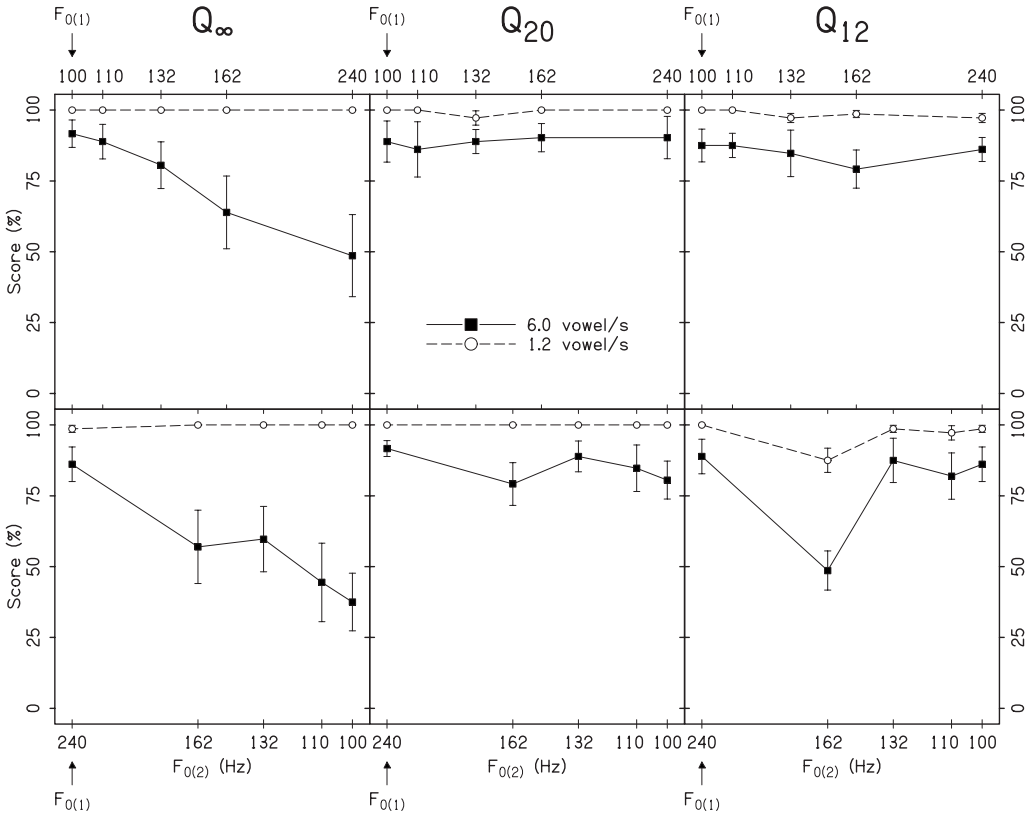


FIG. 2. Shown are group means and standard deviations for a task in which subjects reported the order of six vowels appearing in sequence. Thus, low scores represent a tendency toward segregation. Alternate vowels were at alternate F_0 values ($F_{0(1)}$ and $F_{0(2)}$). In the LowRef conditions (upper panels), $F_{0(1)}$ was fixed at 100 Hz and in the HiRef conditions (lower panels) $F_{0(1)}$ was fixed at 240 Hz. Conditions Q_{20} and Q_{12} involved reduced frequency resolution via 20- and 12-channel noise vocoders. Filled squares represent a Fast condition (6.0 vowel/s) in which streaming can occur, and open circles represent a Slow condition (1.2 vowel/s) that ensures accurate item identification. The abscissa is logarithmic.

Gaudrain *et al.* (2007) attributed low scores in the matched F_0 condition to formant-based streaming. Such a phenomenon has been reported by Dorman *et al.* (1975) with synthesized vowels. Formant-based streaming might be reduced with the recorded vowels used in the current experiment, where small F_0 fluctuations were preserved. Fundamental frequency fluctuations might serve to strengthen the grouping of components comprising individual vowels and limit the grouping of formants across successive vowels, as suggested by Gestalt theory (Bregman, 1990).

In the conditions having spectral degradation (Q_{20} and Q_{12}), the scores are high and do not depend on F_0. Thus, when spectral cues to pitch were reduced in accord with a CI model, F_0-based streaming was reduced or eliminated. Further, these results indicate that the temporal cues to pitch that remained in the vocoded stimuli were not strong enough, in this case, to elicit obligatory streaming. It is potentially interesting to note that these results cannot be explained by a loss of intelligibility since degradation of the stimuli yielded an increase in performance. In addition, vowel identification was confirmed in the Slow condition.

The main finding of this experiment was that no F_0-based streaming appeared when spectral-pitch cues were degraded using a model of a 12- or a 20-channel CI. This result suggests that obligatory streaming is reduced when spectral cues to pitch are reduced in this manner and may not be possible for vowel stimuli based on remaining temporal-pitch cues. This observation is in apparent contrast with studies that observed some streaming in CI recipients (Chatterjee *et al.*, 2006; Hong and Turner, 2006; Cooper and Roberts, 2007). However, these previous observations were generally based on conditions in which some place pitch existed. The current result is also in apparent contrast with the observation of streaming based on temporal pitch in NH (Grimault *et al.*, 2002). One explanation for this discrepancy is that temporal-pitch cues were not sufficiently salient in the noise-band vocoder. This point is addressed in Sec. IV. It is also potentially important that obligatory streaming is strongly influenced by presentation rate (van Noorden, 1975), and that Hong and Turner (2006), Chatterjee *et al.* (2006), and Grimault *et al.* (2002) all used sequences with higher presentation rates (10 stimuli/s) to observe streaming. It then seems plausible that, for an F_0 difference of about one octave, the natural presentation rate used in experiment 1 was not sufficiently high to elicit obligatory streaming under degraded conditions, but that streaming may still be possible. The next experiment assesses this hypothesis.

III. EXPERIMENT 2

As shown by van Noorden (1975), the temporal coherence boundary, the threshold corresponding to obligatory streaming, depends on presentation rate. As shown in experiment 1 of Gaudrain *et al.* (2007), higher presentation rates in

the current paradigm do indeed lead to stronger measures of streaming. Thus, increasing the repetition rate should strengthen the streaming effect and reveal if segregation is possible under the current conditions of severely reduced spectral cues, but preserved temporal cues to pitch. In addition, two envelope cutoff values were employed to more closely examine the role of temporal cues.

A. Materials and method

1. Subjects

Nine subjects aged 18–27 years (mean 21.9) participated. All were native speakers of French and had NH as defined in experiment 1. None of these subjects participated in previous similar experiments, and all were paid an hourly wage for participation.

2. Stimuli

The same six recorded vowels employed in experiment 1 were used. The durations of the vowels were reduced to 133 ms using STRAIGHT. Again, 10 ms ramps were employed. The average F_0's of each vowel were set to 100, 155, and 240 Hz using the same method used in experiment 1. The intact vowels were used for a Q_∞ condition. The same noise-band vocoder used in experiment 1 was again used to process vowels for Q_{20} and Q_{12} conditions. However, unlike experiment 1, two cutoff frequencies (f_c) were used for envelope extraction. A value of f_c=400 Hz was employed to preserve temporal-pitch cues, and a value of f_c=50 Hz was employed to eliminate temporal-pitch cues. As in experiment 1, envelope extraction involved half-wave rectification and eighth order Butterworth lowpass filtering.

The processed vowels were then concatenated to form sequences in the five processing conditions (Q_∞, Q_{20} f_c =400 Hz, Q_{20} f_c=50 Hz, Q_{12} f_c=400 Hz, and Q_{12} f_c =50 Hz). The 36 arrangements having the lowest perceptual distance were selected and divided into three groups having approximately equal mean perceptual distance values. As in experiment 1, these three groups were used for the three F_0 separation conditions. Thus, as in experiment 1, the particular arrangements of vowels were distributed across the F_0 separation conditions, but repeated across the other conditions to ensure that effects associated with the particular order of items were constant.

In this experiment, only the LowRef condition was used, so that $F_{0(1)}$ was always 100 Hz. The low identification score observed for $F_{0(2)}$=162 Hz in the HiRef condition of experiment 1 was hence avoided. As in the first experiment, Slow (1.2 vowel/s) and Fast (7.5 vowel/s) sequences were employed. Slow sequences were repeated 4 times and Fast sequences were repeated 25 times, so that the overall duration in both conditions was 20 s. Stimuli were generated with MATLAB as 16 bit, 44.1 kHz sound files, and were presented as in experiment 1.

3. Procedure

a. Training and selection. Training again began with simple identification of single vowels. Five blocks of 72 vowels were presented. Each block contained four repetitions of each vowel at each $F_{0(2)}$ in a single degradation condition, in random order (4 repetitions $\times$ 6 vowels $\times$ 3 F_0's=72 items). The blocks were presented from the least degraded (Q_∞) to the most degraded (Q_{12} f_c=50 Hz). Visual feedback was provided. Each block was repeated until a score of 96% (69/72) was obtained or a maximum of three repetitions was reached. On average the blocks were repeated 2.0 times for each condition (range 1.7–2.3).

As in experiment 1, training involving the Slow condition sequences followed. The test consisted of seven blocks. Each block was composed of 36 sequences and all the conditions were represented at least twice in random order. Subjects were required to score greater than 95% correct over three successive blocks in each condition to advance to the next stage. One subject was unable to reach the criterion and was dismissed. For seven of the remaining participants, five blocks were sufficient to reach the criterion. The last subject reached the criterion after seven blocks.

b. Streaming test. The test consisted of 5 blocks of 72 sequences each. All conditions (3 $F_{0(2)}$'s, 5 Q's, and 2 Slow/Fast) were represented as equally as possible in each block. For each $F_{0(2)}$, streaming was measured over 12 different arrangements of vowels. Other aspects of the experiment, including the initial 5 s response lockout and the manner of response, were identical to those of experiment 1. Experiment 2 required subjects to participate in three to four 2 h sessions, during which frequent breaks were provided. The experimental procedure was formally approved by a local ethics committee (CPP Sud Est II).

B. Results

Results averaged across subjects are plotted in Fig. 3. As can be seen, scores were uniformly high in the Slow conditions (mean: 98.6% correct), reflecting accurate identification of the constituent items. The subsequent analyses were conducted on the Fast conditions. A two-way ANOVA, using processing condition (Q_∞, Q_{20} f_c=400 Hz, Q_{20} f_c=50 Hz, Q_{12} f_c=400 Hz, and Q_{12} f_c=50 Hz) and F_0 as repeated parameters, showed a significant effect of processing condition [$F(4,28)=23.57$, $p<0.001$] as well as F_0 [$F(2,14)=8.25$, $p<0.01$]. These factors did not interact [$F(8,56)=1.31$, $p=0.26$], indicating that the effect of F_0 was not significantly different between processing conditions. To test for the effect of f_c, separate two-way ANOVAs (using F_0 and f_c as repeated parameters) were performed in each of the degraded conditions. The analyses revealed no effect of f_c in the Q_{12} [$F(1,7)=0.9$, $p=0.37$] or Q_{20} conditions [$F(1,7)=3.86$, $p=0.09$]. The effect of F_0 was significant in the Q_{12} condition [$F(2,14)=7.11$, $p<0.01$], but not in the Q_{20} condition [$F(2,14)=0.47$, $p=0.63$]. The interaction was not significant in the Q_{12} condition [$F(2,14)=0.78$, $p=0.48$], but was at Q_{20} [$F(2,14)=6.37$, $p<0.01$], as suggested by the pattern of results in Fig. 3. An analysis of this interaction using pairwise t-tests revealed no significant difference between f_c values, even at $F_{0(2)}$=100 Hz [$p=0.30$].

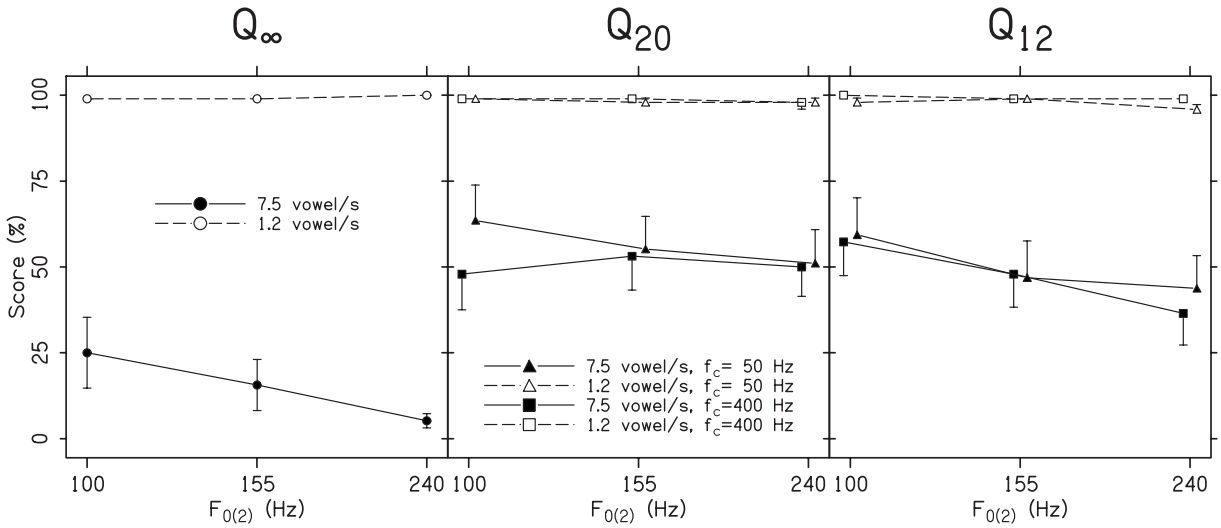


FIG. 3. Shown are group means and standard deviations for the vowel order identification task in experiment 2. Again, low scores represent a tendency toward segregation. $F_{0(1)}$ was at 100 Hz and $F_{0(2)}$ is shown. In each Q condition, scores for the Slow conditions (1.2 vowel/s) are plotted as open symbols, and scores in the Fast conditions (7.5 vowel/s) are plotted as filled symbols. In the conditions Q_{20} and Q_{12} (middle and rightmost panels), scores for conditions in which temporal smoothing (f_c) was 50 Hz are plotted with triangles, and scores for f_c=400 Hz are plotted with squares. The abscissa is logarithmic.

C. Discussion

The scores in the matched F_0 conditions were lower in the current experiment than in experiment 1, indicating that the subjects performed more poorly in this second experiment. This result can likely be attributed to at least two sources. The first is the use of naive listeners in the current experiment and trained listeners in experiment 1. The second is that increasing the presentation rate enhances segregation based on formant structure. Gaudrain *et al.* (2007) argued that vowels having matched F_0 can elicit streaming based on formant structure, as found by Dorman *et al.* (1975). The higher scores in the Q_{20} and Q_{12} conditions support this hypothesis, suggesting, as in Gaudrain *et al.*, 2007, that formant-based streaming is hindered by loss of frequency resolution. Despite the fact that scores were reduced in the $F_{0(2)}$=100 Hz conditions, a significant main effect of $F_{0(2)}$ was nevertheless observed across all conditions. Although this effect was not observed at Q_{20}, it was observed in the isolated Q_{12} condition, indicating that some F_0-based streaming occurred in this degraded condition.

The presence (f_c=400 Hz) or absence (f_c=50 Hz) of temporal-pitch cues did not influence the performance in either degraded condition. The use of an eighth order smoothing filter during envelope extraction ensured that modulation frequencies above these temporal cutoffs were not present at meaningful levels (Healy and Steinbach, 2007). This result suggests that streaming was not based on temporal-pitch cues. Thus, other reasons for F_0-based streaming must be sought.

One obvious F_0-related cue that can be considered is the modulation product. Because of basilar membrane nonlinearity, the periodic modulation in the channels of the vocoder could evoke a component at this modulation frequency and at harmonic multiples. If the envelope modulation periodicity represents the F_0, the modulation product can recreate the original first harmonic. However, when f_c=50 Hz, pitch cues associated with F_0 are removed from the envelope. If amplitude modulation had elicited streaming, a difference should have been observed between the f_c conditions. Thus, although modulation products may be evoked with the present stimuli, they are likely not responsible for the observed pitch-based streaming.

It must then be considered that some remaining spectral cues could be related to the F_0. One such cue involves the first harmonic. The first harmonic can fall either in the first band of the vocoder or below it. It could then possibly provide an F_0-related cue because it can influence the level of the first channel. To test this hypothesis, the level in the first channel was measured in the different conditions. A three-way ANOVA on this measure using F_0, Q, and f_c as repeated parameters, across the six vowels, showed no effect of F_0 [$F(2,10)=0.03$, $p=0.97$] (f_c [$F(1,5)=0.84$, $p=0.40$] and Q [$F(1,5)=449.3$, $p<0.001$]). Thus, the first channel does not appear to contain a consistent F_0-related cue. However, the distribution of harmonics across all channels could have an effect. A Fisher discriminant analysis (FDA) was used to find a linear combination of channel levels that provides the best representation of the F_0 in the Q_{12} condition (using Python MDP, Berkes and Zito, 2007). As shown in Fig. 4, for both f_c=400 and 50 Hz, a linear combination of channel levels can be found to represent the F_0. This linear combination can be used as a metric that represents F_0. In Fig. 4, scores were plotted against this metric to show the relation between this metric and segregation. In conclusion, it is possible that some spectral cues related to the F_0, but not capable of generating a strong pitch sensation, could persist even in the absence of harmonics.

Although the evidence for remaining F_0-related spectral cues in the Q_{12} condition is relatively clear, it remains unclear whether these cues would be present in a real CI and for sounds other than those employed here. It is then important to determine the origin of these cues. The vocoder discards the harmonic structure of the vowels. Hence, if a spectral cue is preserved by the vocoder, it must be encoded by

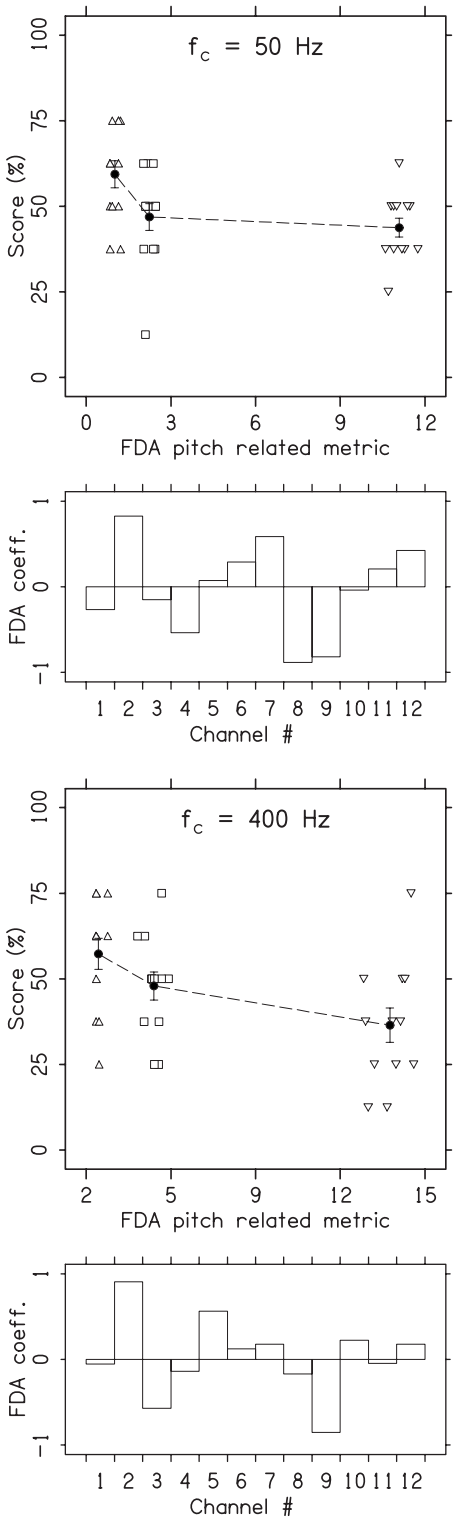


FIG. 4. Upper panels: Scores as a function of the pitch related metric found by FDA on vocoder channel mean levels for vowels processed in the Q_{12} condition, f_c=50 Hz. The upward triangles represent sequences with $F_{0(2)}$ =100 Hz, squares for $F_{0(2)}$=155 Hz, and downward triangles for $F_{0(2)}$ =240 Hz. The coefficients of the linear combination used as a metric are represented in the lower subpanel. Lower panels: Same representation for the Q_{12} condition, f_c=400 Hz.

the spectral envelope that is partially preserved. To examine this, the spectral envelopes of the unprocessed vowels (Q_∞) were extracted using STRAIGHT for the three different F_0's. These envelopes showed slight but consistent differences in the formant regions that are probably attributable to the formant definition in relation to the harmonic density: Higher F_0's tended to produce broader formants. This small broadening effect seems to have been emphasized by the spectral quantization in the 12-channel vocoder.

It would have been interesting to determine whether this cue existed in the Q_{20} condition. However, it was not possible to analyze the Q_{20} condition because FDA requires the number of items (6 vowels $\times$ 3 F_0's=18) to be greater than the number of parameters (20 vocoder bands). It is possible that the greater number of channels reduced the quantization effect, and did not emphasize this F_0-related cue as much as in the 12-channel vocoder.

This analysis of the stimuli suggests that the F_0-related spectral cues are due to the modification of formant definition associated with changes in F_0. Although this phenomenon may occur when perceiving natural vowels uttered at different F_0's, the frequency quantization of the 12-channel noise-band vocoder simply emphasized the effect, as well as making it relatively stochastic across vowels. It is worth noting that, in the current experiment, the formant positions were held constant while changing F_0. In real speech, the formant position and width are related to speaker size, which also drives the nominal F_0 of the speaker. This F_0-related cue may then appear along with some other cues associated with vocal tract length (VTL), which can work to further support streaming. Although streaming in NH listeners seems to be less influenced by changes in VTL than in F_0 (Tsuzaki *et al.*, 2007), this kind of cue could be preserved and used in CI listening.

IV. GENERAL DISCUSSION

The main result of this study is that obligatory streaming was reduced, but still present, with spectral cues reduced in accord with a CI model, but that this streaming was not attributable to temporal-pitch cues. This is consistent with Gaudrain *et al.* (2007) who observed that impoverishing the spectral representation of vowels via simulated broadened auditory tuning hindered pitch-based streaming. Reducing the spectral cues and preserving most of the temporal cues in experiment 1 reduced sufficiently the salience of the two different streams to eliminate obligatory streaming. In the second experiment, emphasizing streaming through a higher presentation rate did not lead to the observation of streaming based on temporal cues, but instead led to the observation of streaming based on a stochastic spectral cue.

The current results are also consistent with previous observations of concurrent speech perception in NH listeners hearing CI simulations. NH listeners are able to take advantage of a pitch difference between a target and a masker to enhance speech recognition in noise (Brokx and Nooteboom, 1982). However, NH listeners are less able to take advantage of pitch differences between speech target and masker when the stimuli are processed with a noise-band vocoder (Stickney *et al.*, 2004). Similarly, Deeks and Carlyon (2004) reported that pulse rate differences did not enhance the reception of concurrent speech in a CI simulation based on unresolved complexes. The absence of pitch-based streaming

observed in experiment 1 suggests that impairment of stream segregation could be partially responsible for low speech perception performance in noise under CI simulation.

However, this result differs from those previously obtained in CI recipients using simple stimuli to test place-pitch or rate-pitch-based streaming (Hong and Turner, 2006; Chatterjee *et al.*, 2006). One explanation for this difference is that the complexity of spectral and temporal cues associated with even simple speech items such as vowels diminishes the availability of F_0 cues following CI processing. Because of the characteristics of speech, F_0 could not be reliably converted into place pitch or rate pitch, such as that associated with the previous experiments. Another possible reason is that the duration of the stimuli employed here was different from that used in previous studies. Hong and Turner (2006) and Chatterjee *et al.* (2006) used 60 and 50 ms stimuli, while the briefest vowels in the current study were 133 ms. A slower presentation rate can prevent obligatory streaming. However, to evaluate if obligatory streaming is involved in concurrent speech segregation in ecological situations, it is important to examine natural speech rates. Because streaming was not observed with speech rates matching those of natural speech in experiment 1 of the current study, it can be concluded that CI users may have difficulty taking advantage of streaming cues in cocktail party situations.

The current CI model employed numbers of channels similar to those found in modern CIs and an overall bandwidth somewhat smaller than that often employed clinically, but more appropriate for vowel reception. Because of this reduced overall bandwidth, and because the number of auditory channels available to a CI user can be somewhat lower than the number of physical electrodes (with the exception of successful "current steering" programming), the spectral detail provided by the current CI model may be even higher than that available in modern CIs. However, the comparison with CI user performance remains difficult because the noise-band vocoder simulation does not exactly mimic the perception of sounds by CI users. There is a long list of patient variables that are associated with CI users, but absent from consideration when using simulations. Also, CI recipients have generally experienced their CI for months prior to experimentation while NH participants in the current experiments were trained on noise-band vocoders for a few hours. Notably, the noise-band vocoder does not exactly mimic stimulation by the CI. A main difference is that the output of the vocoder is acoustic and then subjected to peripheral processing by the ear, while the CI involves direct electric stimulation.

Another effect that potentially reduces the strength of temporal pitch in the vocoder has been suggested by Hanna (1992). In the noise-band vocoder, the analysis filter bank and the resynthesis filter bank are typically the same. Thus, the noise carrier after modulation with the temporal envelope is then filtered again to suppress the sidebands, i.e., the modulation products that fall outside the band. Modulation depth is reduced by this resynthesis filtering for the narrowest bands, and temporal-pitch cues are then weakened. To fully preserve the voicing modulation, the bandwidth must be greater than twice the F_0. Using this metric, temporal-pitch cues were intact in the current vocoder channels beyond No. 6 in the Q_{12} condition and beyond No. 14 in the Q_{20} condition. In addition, as described by Hanna (1992), the peripheral filters of the normal ear play a role similar to that of the resynthesis filters of the vocoder, weakening again the modulation depth. Thus, although subjects were probably able to perceive temporal-pitch cues, their depth was reduced in the lower bands of the vocoder. In CI processors, neither resynthesis nor peripheral filtering occurs and temporal-pitch cues are not degraded in this way. Indeed, Laneau *et al.* (2006) observed that CI users had better F_0 discrimination abilities than NH listeners hearing noise-band vocoders.

Although it was found that temporal-pitch cues were not sufficient to produce obligatory streaming of vowel sequences, it cannot be entirely ruled out that pitch-based obligatory streaming can occur in CI users. In particular, it was found that an F_0-related spectral cue may have induced obligatory streaming. This spectral cue is potentially related to formant definition and was then not present in simpler stimulation. The current findings then suggest that these cues could potentially be available to CI listeners. Many attempts have been made in the past few years to provide a better encoding of pitch for CI recipients. Increasing the number of bands in the low frequencies better captures the first harmonic and can improve the perception of pitch (Carroll and Zeng, 2007). Unfortunately, with a fixed number of channels, increasing the number of bands in the low frequency region leads to a decrease in the number of bands in the higher regions, which appears to be detrimental for speech intelligibility. Hence, there seems to be a trade-off between pitch perception and speech intelligibility. Moreover, despite the increasing number of channels in CIs (up to 22), it seems that most CI recipients do not show a speech reception benefit from more than seven bands (Friesen *et al.*, 2001). These results suggest that pitch cues should be enhanced in existing bands to avoid the degradation of spectral cues required for speech perception. Instead of increasing spectral-pitch cues, Green *et al.* (2005) have enhanced temporal-pitch cues by adding to standard processing 100% AM at the F_0 frequency in all channels This manipulation improved the perception of prosody. However, again, it appeared that the modified processing had a detrimental effect on vowel recognition. These two strategies to enhance pitch perception in CI users do not account for the F_0-related spectral cues found in the current study. Further investigation is then required to evaluate the ability of CI users to take advantage of these cues for segregation of speech in ecological situations.

V. CONCLUSIONS

(1) Temporal-pitch cues available from noise-band vocoder simulations of a CI are not sufficient to induce obligatory streaming of speech materials at realistic speech rates.
(2) In contrast, the quantization of the spectrum in the vocoder enhances an F_0-related spectral cue that is capable of inducing streaming. This cue might be available to CI users.
(3) The use of temporal periodicity cues to induce obliga-

tory streaming in CI users is unlikely, unless it is assumed that these cues are stronger in actual CIs than in CI simulations.

ACKNOWLEDGMENTS

The authors wish to thank Andrew J. Oxenham and Christophe Micheyl for their helpful comments on this study. The authors would also like to thank two anonymous reviewers for helpful comments on a previous version of this manuscript. This work was supported in part by NIH Grant No. DC08594, by a doctoral grant from the Région Rhône-Alpes (France), and by Grant No. JC6007 from the Ministère de l'Enseignement Supérieur et de la Recherche (France). Manuscript preparation was supported in part by UK Medical Research Council Grant No. G9900369.

[1]Perceptual distance between vowels was calculated as the weighted Euclidian distance (in barks) in the F1–F'2 space, where F1 is the frequency of the first formant and F'2 is the frequency of the *effective second formant* defined as the weighted sum of the second to fourth formants (de Boer, 2000). Perceptual distance for a given sequence was then the sum of the perceptual distances between consecutive vowels.

American National Standards Institute (**1995**). *ANSI S3.7-R2003: Methods for Coupler Calibration of Earphones* (American National Standard Institute, New York).

American National Standards Institute (**2004**). *ANSI S3.21-2004: Methods for Manual Pure-Tone Threshold Audiometry* (American National Standard Institute, New York).

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1241.

Beauvois, M. W., and Meddis, R. (**1996**). "Computer simulation of auditory stream segregation in alternating-tone sequences," J. Acoust. Soc. Am. **99**, 2270–2280.

Berkes, P., and Zito, T. (**2007**). "Modular toolkit for data processing, version 2.1," http://mdp-toolkit.sourceforge.net

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Bregman, A. S. (**1978**). "Auditory streaming is cumulative," J. Exp. Psychol. Hum. Percept. Perform. **4**, 380–387.

Bregman, A. S., and Campbell, J. (**1971**). "Primary auditory stream segregation and perception of order in rapid sequences of tones," J. Exp. Psychol. **89**, 244–249.

de Boer, B. (**2000**). "Self-organization in vowel systems," J. Phonetics **28**, 441–465.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Burns, E. M., and Viemeister, N. F. (**1981**). "Played-again SAM: Further observations on the pitch of amplitude-modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.

Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Carlyon, R. P., Long, C. J., Deeks, J. M., and McKay, C. M. (**2007**). "Concurrent sound segregation in electric and acoustic hearing," J. Assoc. Res. Otolaryngol. **8**, 119–133.

Carroll, J., and Zeng, F. (**2007**). "Fundamental frequency discrimination and speech perception in noise in cochlear implant simulations," Hear. Res. **231**, 42–53.

Chatterjee, M., Sarampalis, A., and Oba, S. I. (**2006**). "Auditory stream segregation with cochlear implants: A preliminary report," Hear. Res. **222**, 100–107.

Cooper, H. R., and Roberts, B. (**2007**). "Auditory stream segregation of tone sequences in cochlear implant listeners," Hear. Res. **225**, 11–24.

Deeks, J. M., and Carlyon, R. P. (**2004**). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation," J. Acoust. Soc. Am. **115**, 1736–1746.

Dorman, M. F., Cutting, J. E., and Raphael, L. J. (**1975**). "Perception of temporal order in vowel sequences with and without formant transitions," J. Exp. Psychol. Hum. Percept. Perform. **104**, 147–153.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2410.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.

Dudley, H. (**1939**). "The automatic synthesis of speech," Proc. Natl. Acad. Sci. U.S.A. **25**, 377–383.

Elhilali, M., and Shamma, S. (**2007**). in *Hearing: From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer-Verlag, Berlin).

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Gaudrain, E., Grimault, N., Healy, E. W., and Béra, J.-C. (**2007**). "Effect of spectral smearing on the perceptual segregation of vowel sequences," Hear. Res. **231**, 32–41.

Geurts, L., and Wouters, J. (**2001**). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **109**, 713–726.

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.

Grimault, N., Bacon, S. P., and Micheyl, C. (**2002**). "Auditory stream segregation on the basis of amplitude modulation rate," J. Acoust. Soc. Am. **111**, 1340–1348.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (**2001**). "Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing," Br. J. Audiol. **35**, 173–182.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (**2000**). "Influence of peripheral resolvability on the perceptual segregation of harmonic tones differing in fundamental frequency," J. Acoust. Soc. Am. **108**, 263–271.

Grose, J. H., and Hall, J. W. (**1996**). "Perceptual organization of sequential stimuli in listeners with cochlear hearing loss," J. Speech Hear. Res. **39**, 1149–1158.

Hanna, T. E. (**1992**). "Discrimination and identification of modulation rate using a noise carrier," J. Acoust. Soc. Am. **91**, 2122–2128.

Hartmann, W. M., and Johnson, D. (**1991**). "Stream segregation and peripheral channeling," Music Percept. **9**, 115–184.

Healy, E. W., and Bacon, S. P. (**2002**). "Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing," J. Speech Lang. Hear. Res. **45**, 1262–1275.

Healy, E. W., and Steinbach, H. M. (**2007**). "The effect of smoothing filter slope and spectral frequency on temporal speech information," J. Acoust. Soc. Am. **121**, 1177–1181.

Hong, R. S., and Turner, C. W. (**2006**). "Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients," J. Acoust. Soc. Am. **120**, 360–374.

Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, 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.

Laneau, J., Moonen, M., and Wouters, J. (**2006**). "Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants," J. Acoust. Soc. Am. **119**, 491–506.

Mackersie, C., Prida, T., and Stiles, D. (**2001**). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **44**, 19–28.

McCabe, S. L., and Denham, M. J. (**1997**). "A model of auditory streaming," J. Acoust. Soc. Am. **101**, 1611.

Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).

Moore, B. C. J., and Carlyon, R. P. (**2005**). in *Pitch: Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York).

Moore, B.C.J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta. Acust. Acust. **88**, 320–332.

Moore, B. C. J., and Peters, R. W. (**1992**). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency

selectivity," J. Acoust. Soc. Am. **91**, 2881–2893.
Nooteboom, S. G., Brokx, J. P. L., and de Rooij, J. J. (**1978**). in *Studies in the Perception of Language*, edited by W. J. M. Level and G. B. F. d'Arcais (Wiley, New York), pp. 75–107.
Patel, A. D., Iversen, J. R., and Rosenberg, J. C. (**2006**). "Comparing the rhythm and melody of speech and music: The case of British English and French," J. Acoust. Soc. Am. **119**, 3034–3047.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," J. Acoust. Soc. Am. **112**, 2074–2085.
Rogers, C. F., Healy, E. W., and Montgomery, A. A. (**2006**). "Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions," J. Acoust. Soc. Am. **119**, 2276–2287.
Rose, M. M., and Moore, B. C. J. (**2005**). "The relationship between stream segregation and frequency discrimination in normally hearing and hearing-impaired subjects," Hear. Res. **204**, 16–28.
Rose, M. M., and Moore, B. C. J. (**1997**). "Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **102**, 1768–1778.
Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I: Basic psychophysics," Hear. Res. **11**, 157–189.
Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Stainsby, T. H., Moore, B. C. J., and Glasberg, B. R. (**2004**). "Auditory streaming based on temporal structure in hearing-impaired listeners," Hear. Res. **192**, 119–130.
Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.
Stickney, G. S., Zeng, F., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Tong, Y. C., and Clark, G. M. (**1985**). "Absolute identification of electric pulse rates and electrode positions by cochlear implant patients," J. Acoust. Soc. Am. **77**, 1881–1888.
Townshend, B., Cotter, N., Compernolle, D. V., and White, R. L. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–115.
Tsuzaki, M., Takeshima, C., Irino, T., and Patterson, R. D. (**2007**). in *Hearing: From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V.Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer-Verlag, Berlin).
van Noorden, L. P. A. S. (**1975**). "Temporal coherence in the perception of tones sequences," Ph.D. thesis, Eindhoven University of Technology, The Netherlands.
Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (**1999**). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," J. Acoust. Soc. Am. **106**, 938–945.
Vliegen, J., and Oxenham, A. J. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am. **105**, 339–346.
Zeng, F. G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.

Discrimination of interaural temporal disparities conveyed by high-frequency sinusoidally amplitude-modulated tones and high-frequency transposed tones: Effects of spectrally flanking noises

Leslie R. Bernstein and Constantine Trahiotis
Department of Neuroscience and Department of Surgery (Otolaryngology), University of Connecticut Health Center, Farmington, Connecticut 06030

(Received 7 February 2008; revised 31 July 2008; accepted 4 August 2008)

Dreyer and Oxenham [(2008). "Effects of level and background noise on interaural time difference discrimination for transposed stimuli," J. Acoust. Soc. Am. **123**, EL1–EL7] reported that spectrally flanking noise increased threshold interaural temporal disparities (ITDs) conveyed by high-frequency transposed tones but rendered them indiscriminable when they were conveyed by high-frequency sinusoidally amplitude-modulated (SAM) tones. This study extends those observations and evaluates the role of "off-frequency listening." Threshold ITDs were measured using 4-kHz-centered transposed or SAM tonal "targets." In "baseline" conditions, targets were presented without spectrally flanking noise. Additionally, targets were presented along with continuous diotic broadband Gaussian noise spectrally "notched" between 3.6 and 4.4 kHz. In another condition, only the high-pass segment of the notched noise was continuously present. In the final condition, only the low-pass segment was continuously present. Results indicate that (1) relative to baseline, adding notched noise resulted in similar relative increases of threshold ITDs for both SAM and transposed targets; (2) the presence of the high-pass segment of the notched noise resulted in greater relative increases in threshold ITDs over those obtained in baseline conditions for SAM tones as compared to transposed tones; (3) comparisons among all of the data were consistent with the interpretation that both on-frequency and off-frequency processing of envelope-based ITDs can be disrupted by the presence of a notched noise. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2980523]

PACS number(s): 43.66.Pn, 43.66.Ba [RLF] Pages: 3088–3094

I. INTRODUCTION

Dreyer and Oxenham (2008) recently reported data concerning the discriminability of interaural temporal disparities (ITDs) conveyed by the envelopes of high-frequency stimuli.[1] The purpose of their study was to determine if the resolution of envelope-based ITDs depends upon listeners' use of "off-frequency" information (i.e., information in spectral regions surrounding those "tuned" to the center frequency of the stimulus). To that end, Dreyer and Oxenham (2008) measured threshold ITDs using a two-alternative forced-choice adaptive procedure and two classes of high-frequency complex waveforms: sinusoidally amplitude-modulated (SAM) tones and "transposed" tones. Transposed tones are specially constructed waveforms designed such that their envelopes convey temporal information to the high-frequency channels of the auditory system that is similar to that conveyed by low-frequency tones to the low-frequency channels of the auditory system (e.g., van de Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2002).

Dreyer and Oxenham's (2008) general strategy was to evaluate off-frequency listening by comparing threshold ITDs obtained when 4-kHz-centered targets were presented along with a diotic broadband noise containing a spectral "notch" between 3.6 and 4.4 kHz with threshold ITDs obtained in the absence of such noise. They reported two interesting and important results. First, threshold ITDs measured with transposed tones in the presence of the notched noise were elevated by a factor of two or so as compared to those measured in the absence of the surrounding noise. Second, for the SAM tones, adding the notched noise resulted in the listeners being unable to perform the task.

Being intrigued and perplexed by both of these findings, we conducted pilot studies with similar stimuli. In contrast to Dreyer and Oxenham's (2008) findings, we found that the presence of the notched noise did not make the task impossible when the ITDs were conveyed by the envelopes of SAM tones. Rather, our results indicated that the presence of a diotic notched noise resulted in similar increases in threshold ITD for *both* transposed and SAM tones. Motivated by these informal findings, we decided to conduct a formal experiment. The goal was to obtain a set of data that would, in their totality, help to assess the effects of notched noise when listeners attempt to resolve envelope-based ITDs that are conveyed within high-frequency channels and to help to evaluate whether off-frequency listening is involved.

Two different types of explanation of Dreyer and Oxenham's (2008) and our preliminary findings occurred to us. The first explanation relies on "upward spread of masking" or "upward spread of excitation" (see Wegel and Lane, 1924; Egan and Hake, 1950; O'Malley and Feth, 1979; Schroeder *et al.*, 1979 for relevant psychophysical data and Kim and

 © 2008 Acoustical Society of America

Molnar, 1979 for relevant neurophysiological data). Based on upward spread, spectral regions of the cochlear partition tuned to frequencies *above* 4 kHz would be expected to convey salient ITD information that could enhance discrimination of ITDs by augmenting the ITD information carried by the 4 kHz "target" region (see van de Par *et al.*, 2000). If that were the case, then the high-pass segment of the notched noise could adversely affect performance by disrupting the use of ITD information conveyed by those higher-frequency regions. There is also a manner in which upward spread of excitation could degrade performance. Specifically, the low-pass segment of the notched noise could disrupt ITD information conveyed by the 4 kHz target region and, perhaps to some extent, ITD information conveyed by regions of the cochlea tuned to frequencies above 4 kHz.

It is also logically possible that *downward* spread of excitation and downward masking could play a role. Assuming downward spread of excitation, then the presence of the low-pass segment of the notched noise could degrade sensitivity to ITD by preventing listeners from utilizing off-frequency ITD information conveyed by spectral regions tuned *below* the 4-kHz target region. Again, assuming downward spread, the high-pass segment of the notched noise could disrupt ITD information conveyed by the 4-kHz target region and, perhaps to some extent, ITD information conveyed by spectral regions below 4-kHz.

We believe, however, that upward spread would affect ITD processing more than would downward spread. This belief is based on two different types of psychophysical results. First, a variety of modern auditory masking studies (some of which are cited above) are consistent with the classic findings of Wegel and Lane (1924) and have demonstrated that masking is highly asymmetric in that a given masker produces a greater amount of masking for targets above its spectral region than for targets below its spectral region. For example, Schroeder *et al.* (1979) demonstrated that the masked threshold of a critical-band-wide noise centered at 1 kHz dropped at a substantially greater rate when the frequency of the 80 dB SPL tonal masker was raised above 1 kHz than when it was reduced below 1 kHz. Second, van de Par *et al.* (2000), who also studied the use of off-frequency binaural information at high frequencies, found that it was off-frequency information in spectral regions above the signal that was salient for the discrimination of high-frequency envelope-based ITDs.

II. EXPERIMENT

In order to evaluate these possibilities, threshold ITDs were measured using SAM or transposed tonal "targets" centered at 4 kHz. All targets were 100% modulated at 128 Hz. Four general stimulus conditions were employed. In the first, or "baseline" condition, the targets were presented without spectrally flanking noise. In the second condition, the targets were presented along with a continuous diotic broadband (10 kHz, low-pass) Gaussian noise having a spectral "notch" between 3.6 and 4.4 kHz. This stimulus condition is like the notched noise condition employed by Dreyer and Oxenham (2008). In the third condition, only the high-pass segment (4.4–10 kHz) of the notched noise was continuously present. In the fourth condition, only the low-pass segment (3.6 kHz low-pass) of the notched noise was continuously present.

A. Procedure

The SAM and transposed-tone targets were generated digitally using a sampling rate of 20 kHz (TDT AP2) and were low-pass filtered at 8.5 kHz (TDT FLT2). They were presented at 55, 75, or 85 dB SPL. The duration of the targets was 300 ms including 20-ms $\cos^2$ rise-decay ramps. The spectrum levels of the notched, 4.4 kHz high-pass, and 3.6 kHz low-pass noises were each 35 dB below the respective level of the target (i.e., 20, 40, or 50 dB). This relative spectrum level was chosen to ensure that we employed a relative spectrum level that was no less than that used by Dreyer and Oxenham (2008). All of the noises, which were presented continuously during the experiment, were generated "offline" in 1 h segments using commercial software (ADOBE AUDITION 1.5©) and were stored on compact disks. During the experiment, the noises were appropriately mixed with the target stimuli and presented to the listeners via TDH-39 earphones. When the baseline (no noise flanking the target) and 4.4-kHz high-pass noise conditions were employed, a continuous diotic noise, low passed at 1.3 kHz (with a spectrum level −35 dB *re* the level of the target), was added to preclude listeners' use of low-frequency distortion products arising from normal nonlinear peripheral auditory processing (e.g., Nuetzel and Hafter, 1976; Bernstein and Trahiotis, 1994). Dreyer and Oxenham (2008) employed a diotic noise low-passed at 400 Hz toward a similar end.

Threshold ITDs were measured using a two-cue, two-alternative, forced-choice, adaptive task. Each trial consisted of a warning interval (500 ms) and four 300-ms observation intervals separated by 400 ms. Each interval was marked visually by a computer monitor. Feedback was provided for approximately 400 ms after the listener responded. The stimuli in the first and fourth intervals were diotic. The listener's task was to detect the presence of an ongoing ITD (left-ear leading) that was presented with equal *a priori* probability in either the second or the third interval. The remaining interval, like the first and fourth intervals, contained diotic stimuli. Ongoing ITDs were imposed by applying linear phase-shifts to the representation of the signals in the frequency domain and then gating the signals destined for the left and right ears coincidentally, after transformation to the time domain. The starting phases of the envelopes and carriers of the targets were chosen randomly for each observation interval both within and across trials. The ITD for a particular trial was determined adaptively in order to estimate 70.7% correct (Levitt, 1971). The initial step size for the adaptive track corresponded to a factor of 1.584 (equivalent to a 2-dB change of ITD) and was reduced to a factor of 1.122 (equivalent to a 0.5 dB change of ITD) after two reversals. A run was terminated after 12 reversals and threshold was defined as the geometric mean of the ITD across the last 10 reversals.

Four normal-hearing adults served as listeners. Two of the listeners (BT and KM) were highly sensitive to ITDs in

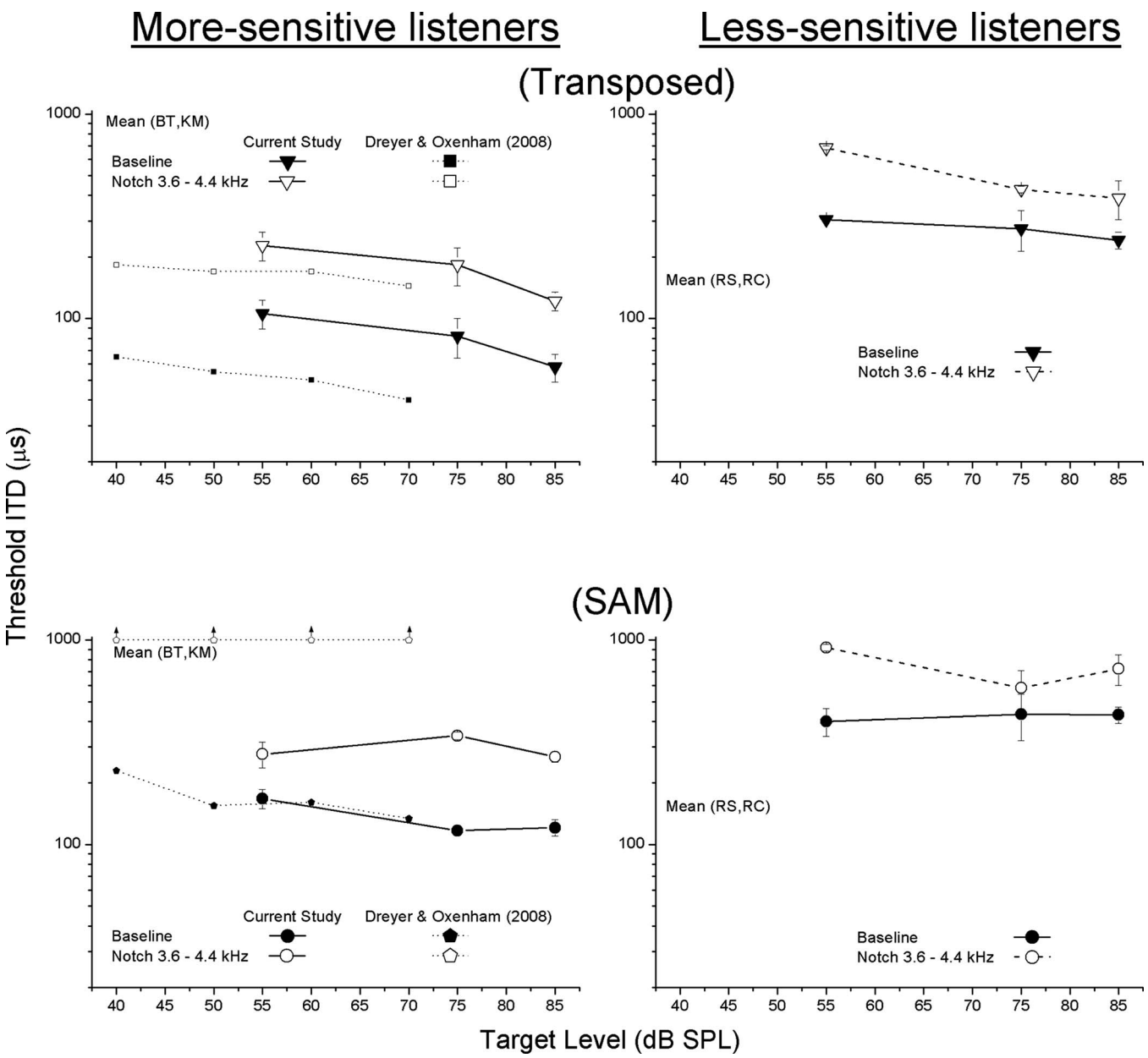


FIG. 1. Threshold ITDs measured as a function of the overall level of the target. Left and right panels depict average threshold ITDs obtained from the two more-sensitive and the two less-sensitive listeners, respectively. Panels in the top and bottom rows depict data obtained with the 4-kHz-centered transposed and SAM targets, respectively. The rate of modulation was 128 Hz. Closed symbols represent the data obtained in the baseline condition; open symbols represent the data obtained when the targets were surrounded by a 3.6–4.4 kHz notched noise. The error bars represent ±1 standard error of the mean. The small squares in the left panel represent average threshold ITDs transcribed from Dreyer and Oxenham (2008).

the baseline condition and two others (RS and RC) were less sensitive. By comparing data obtained across both pairs of listeners, it was possible to evaluate potential interactions between the effects produced by noise flanking the targets and baseline sensitivity to envelope-based ITDs. Particular stimulus combinations were chosen pseudorandomly and three consecutive estimates of threshold were obtained for each of the 24 stimulus combinations (two types of target ×four noise conditions×three levels) before moving onto the next one. Twelve estimates of threshold were obtained for each stimulus combination and the final values of threshold for each listener and stimulus combination were obtained by computing the median of those 12 estimates.

B. Results and discussion

1. Baseline and notched noise conditions

The two plots in the top row of Fig. 1 depict threshold ITDs measured as a function of overall level for the transposed targets. The left and right panels depict average threshold ITDs obtained from our two more-sensitive and our two less-sensitive listeners, respectively. Closed symbols represent the data obtained in the baseline condition; open symbols represent the data obtained when the targets were surrounded by a 3.6–4.4 kHz notched noise. The error bars represent ±1 standard error of the mean. The small squares in the left panel represent average threshold ITDs transcribed from Dreyer and Oxenham (2008).[2] In the same fashion, the

two plots in the bottom row of Fig. 1 depict the threshold ITDs obtained when the targets were SAM tones.

The data obtained in both studies show that (1) the presence of the notched noise elevated thresholds over those obtained in the corresponding baseline conditions and (2) there was little or no effect of varying the overall level of the target in either the baseline or the notched noise condition. Visual comparisons among the threshold ITDs for transposed stimuli (top row, left-hand panel) reveal that Dreyer and Oxenham's (2008) listeners were substantially more sensitive than our more-sensitive listeners in the baseline condition but were only slightly more sensitive in the notched noise condition. The right-hand panel in the top row shows that the threshold ITDs obtained from our two less-sensitive listeners in the baseline condition are about a factor of three greater than those obtained from our more-sensitive listeners. That difference notwithstanding the presence of the notched noise elevated thresholds by about a factor of two to three over those measured in the baseline condition, just as was the case for our more-sensitive listeners.

We now turn to the data collected with SAM stimuli (bottom row). In the baseline condition, threshold ITDs obtained with SAM stimuli were very similar for Dreyer and Oxenham's (2008) listeners and for our more-sensitive listeners (left-hand panel). As was the case for the transposed stimuli (top row), our less-sensitive listeners' thresholds for SAM stimuli (right-hand panel) were approximately three times those of the more-sensitive listeners. In the presence of the notched noise, both our more-sensitive and our less-sensitive listeners required a doubling to tripling of ITD in order to reach threshold, just as was found with the transposed stimuli. This outcome differs greatly from Dreyer and Oxenham's (2008) as their listeners were unable to discriminate ITDs conveyed by SAM tones in the presence of the notched noise (indicated by the upward arrows atop the squares in the left-hand panel).

The data in Fig. 1 were subjected to a three-factor (two noise conditions × three overall levels × two stimulus types), within subjects analysis of variance. The error terms for the main effects and for the interactions were the interaction of the particular main effect (or the particular interaction) with the subject "factor" (Keppel, 1973). Consistent with visual inspection of the data: (1) the main effect of noise condition (presence or absence of the notched noise) was significant (assuming an α of 0.05) [$F(1,3)=18.5$, $p=0.02$]; (2) the main effect of stimulus type (SAM or transposed) was also significant [$F(1,3)=15.0$, $p=0.03$]; (3) neither the main effect of overall level nor any of the interactions among the factors was significant.

Thus, the results of our formal experiment are in accord with the results of the pilot experiment that motivated this study in that threshold ITDs are elevated relatively similarly for both high-frequency SAM and transposed tones. Potential factors that might account for the discrepancy between Dreyer and Oxenham's (2008) and our results will be discussed in Sec. III.

2. Potential role of off-frequency listening

Each row of panels in Fig. 2 displays data obtained from a pair of listeners. The particular pairings were chosen because they reflect a similar patterning of the data across conditions. As it turned out, each such pairing (BT/RS and KM/RC) consists of one of the more-sensitive listeners (BT or KM) and one of the less-sensitive listeners (RS or RC). The bars within each plot represent, for each listener, threshold ITDs normalized against the mean threshold ITD obtained in the respective SAM (left-hand panels) or transposed (right-hand panels) target baseline conditions. The data were collapsed across all three levels of presentation of the targets (55, 75, and 85 dB SPL). The error bars atop each bar within each plot represent the standard error of the mean calculated across all three levels of the stimuli. The small size of the error bars reflects the finding that the relative effects found with the low-pass and high-pass segments, like those analyzed statistically for the notched noise condition (Fig. 2), also did not vary appreciably with level.

Comparisons among the threshold ITDs obtained from the four listeners in the conditions shown in Fig. 2 were evaluated via one-tailed paired t-tests utilizing an alpha of 0.05.

a. High-pass segment of the notched noise. Recall that, because of upward spread of excitation of the energy of the target, spectral regions of the cochlea tuned above 4 kHz could carry salient ITD information that could augment ITD information carried by the 4-kHz region. If that were the case, then the high-pass segment of the notched noise would adversely affect performance by disrupting the use of such information. In fact, for SAM targets, the data from all four listeners indicate that the presence of the 4.4-kHz high-pass noise increased threshold ITDs by a factor of about 1.5–2.0 over those measured in the baseline conditions. Across the four listeners, thresholds ITDs obtained with the high-pass segment were significantly larger than their baseline counterparts for SAM tones ($t=5.6$, $df=11$, $p<0.0005$). To the degree that upward spread of excitation of the target information plays a greater role than downward spread of excitation of the high-pass segment of the noise into the target region, this outcome is consistent with the listeners' use of an off-frequency listening strategy in the SAM baseline condition.

When the targets were transposed tones, however, the threshold ITDs obtained from listeners BT and RC were essentially equal to those obtained in the baseline condition while those obtained from listeners RS and KM were elevated by about a factor of 1.25 over those obtained in the baseline condition. In this case, across the four listeners, thresholds ITDs obtained with the high-pass segment were not significantly larger than their baseline counterparts ($t=1.3$, $df=11$, $p=0.11$). This lack of statistical significance could reflect (1) that the two pairs of listeners employed different strategies; (2) that, overall, there truly is no substantial effect produced by the addition of the high-pass segment when transposed tones serve as targets; or (3) that there is an effect produced by the high-pass segment but data from only four listeners do not yield sufficient statistical power to reveal it.

These statistical outcomes notwithstanding, visual inspection suggests that the relative increases in threshold

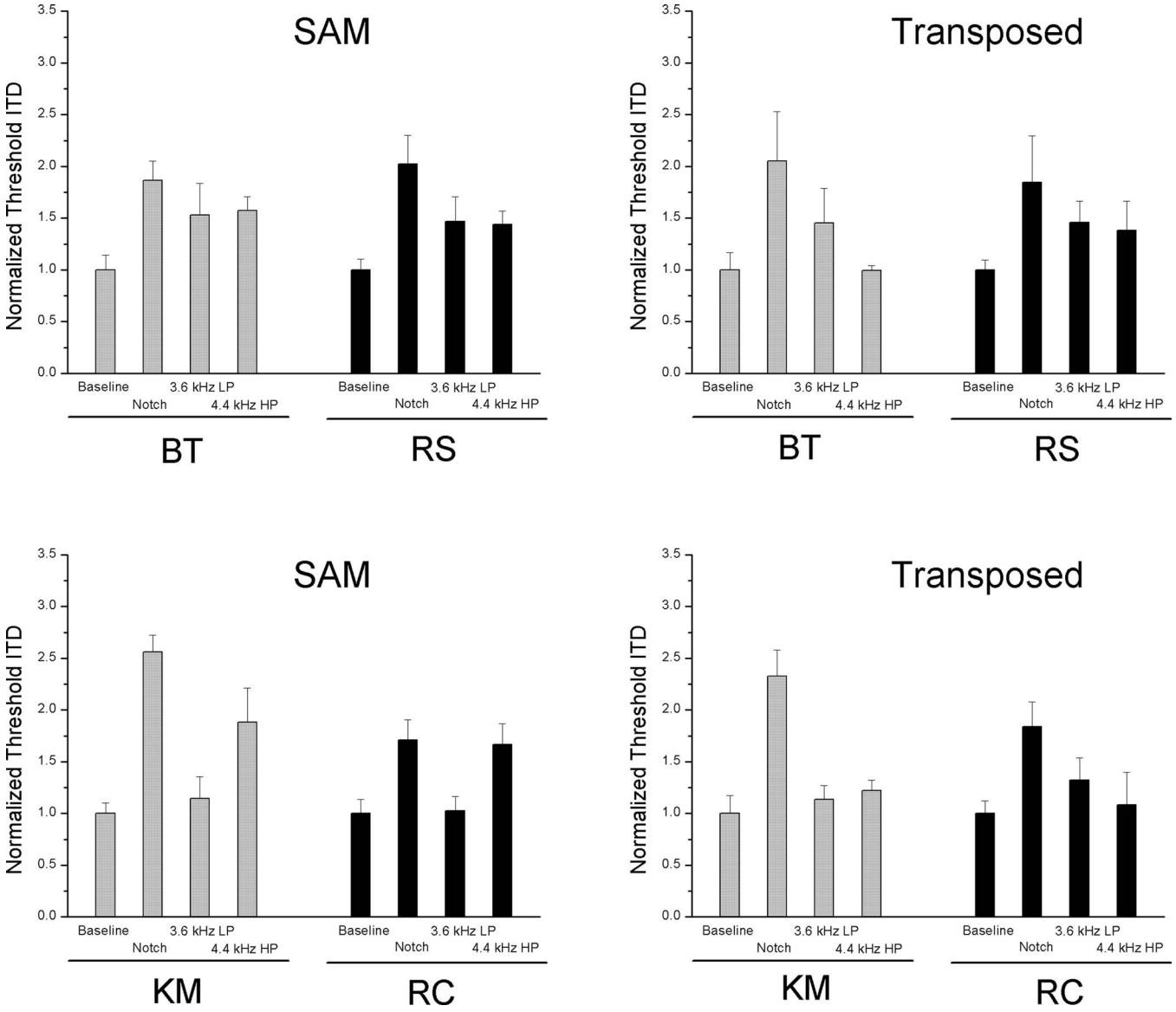


FIG. 2. Each row displays data obtained from a pair of listeners (BT/RS or KM/RC). The bars within each plot represent, for each listener, threshold ITDs normalized against the mean threshold ITD obtained in the respective SAM (left-hand panels) or transposed (right-hand panels) target baseline conditions. The data were collapsed across all three levels of presentation of the targets (55, 75, and 85 dB SPL). Error bars represent one standard error of the mean.

ITDs as compared to baseline in the presence of the high-pass segment that did occur were larger for SAM tones than for transposed tones. Those differences were found to be statistically significant via a two-tailed t-test ($t=4.1$, $df=11$, $p=0.002$). This finding suggests that ITD information conveyed by spectral regions tuned above the center frequency of the target is not necessary for a listener to show enhanced processing of ITDs found with high-frequency transposed stimuli versus SAM tones in the baseline conditions.

b. Low-pass segment of the notched noise. Under the assumption that upward spread of excitation is more salient than is downward spread, the evaluation of deleterious effects attributable to only the low-pass segment of the notched noise is a bit more complicated than was the case for the high-pass segment. This is so because presenting only the low-pass segment of the notched noise along with the target could allow the listener to use "off-frequency" ITD information within cochlear regions tuned well above 4 kHz. That listening strategy would diminish, if not eliminate, deleterious effects on ITD processing stemming from upward spread of excitation of the low-pass segment's energy into the 4-kHz spectral region. Of course, such a strategy could not be usefully employed in the notched noise condition because of the presence of the high-pass segment. Therefore, increases in threshold ITD over the baseline condition produced by the presence of only the low-pass segment of the notched noise are likely to represent underestimates of the deleterious effects produced by that segment when it is accompanied by the high-pass segment in the notched noise condition.

Returning to the data in Fig. 2 with these notions in mind, note that, for SAM tones, listener BT's and listener RS's threshold ITDs increased by a factor of about 1.5 over those measured in the baseline condition when the targets were presented along with the 3.6-kHz low-pass segment. In contrast, threshold ITDs for listeners KM and RC did not increase. These differential outcomes notwithstanding, across the four listeners, thresholds ITDs obtained with the low-

pass segment were significantly larger than their baseline counterparts ($t=2.2$, $df=11$, $p=0.024$).

The data are consistent with the hypothesis that listeners KM and RC were better able to utilize an off-frequency listening strategy than were listeners BT and RS in order to "escape" deleterious effects from upward spread of excitation of the low-pass segment of the notched noise into the 4-kHz target region. In order for that hypothesis to be valid, it would be necessary to show that the low-pass segment of the notched noise, in and of itself, degrades sensitivity to ITDs. The fact that threshold ITDs measured in the notched noise condition are significantly larger than those measured in the presence of either its low-pass ($t=3.6$, $df=11$, $p=0.002$) or its high-pass segment ($t=2.2$, $df=11$, $p=0.027$) provides the required support. This is so because if the low-pass segment of the notched noise had no effect, then threshold ITDs obtained with the notched noise and with its high-pass segment would be equivalent. Instead, the notched noise increased threshold ITDs significantly more so than did only its high-pass segment.

The data in Fig. 2 indicate that, for transposed tones, the presence of the low-pass segment of the notched noise elevated threshold ITDs for listeners BT, RS, and RC by factors of 1.45, 1.46, and 1.32, respectively, over those measured in the baseline condition. For listener KM, threshold ITD increased by only 1.13. As was true for SAM tones, across the four listeners, thresholds ITDs obtained with the low-pass segment were significantly larger than their baseline counterparts ($t=2.1$, $df=11$, $p=0.028$). Also, as was true for SAM tones, threshold ITDs obtained with the notched noise were, for all four listeners, significantly larger than those obtained in the presence of either the low-pass ($t=2.4$, $df=11$, $p=0.016$) or the high-pass segments ($t=5.4$, $df=11$, $p<0.0005$) in isolation. Thus, following the same lines of reasoning presented above for SAM tones, the data are consistent with the hypothesis that all four listeners employed an off-frequency listening strategy to escape deleterious effects of the low-pass segment when it was presented in isolation.

III. GENERAL DISCUSSION

As did Dreyer and Oxenham (2008), we found that the presence of a notched noise that spectrally flanked the high-frequency spectral region conveying ITD information resulted in increases in threshold ITD as compared to baseline conditions. In addition, we were able to confirm their finding that threshold ITDs were essentially unaffected by changes in the overall level of the stimuli over a 30 dB range. On the other hand, some of the results of this investigation and the conclusions based on them differ from some of those of Dreyer and Oxenham (2008).

The most striking difference between the studies is that, while none of Dreyer and Oxenham's (2008) listeners was able to perform the ITD-discrimination task with SAM tones in the presence of notched noise, all of our listeners (i.e., those in the formal study and those in the pilot study) were able to do so. In fact, we found that the notched noise produced about the same relative increase in ITD thresholds for both SAM and transposed tones over their respective baseline conditions. In attempts to reconcile the differences between the two studies, we collected ancillary data using a two-interval forced-choice procedure like that employed by Dreyer and Oxenham (2008) and a pulsed notched noise which, like theirs, was gated on 400 ms before the first observation interval and gated off 200 ms after the second observation interval that composed a trial. The data in Fig. 1 were validated in that, even with these stimulus conditions and with no additional "practice," all four of our listeners performed the task with both SAM and transposed tones and the group's threshold ITDs were not appreciably different from those reported within the main experiments. In a final manipulation, the spectrum level of the notched noise was increased by 10 dB while leaving unaltered, at 75 dB SPL, the levels of the SAM and transposed targets. Three of the four listeners participated in this follow-up study. All three of them were unable to perform the task (ITDs less than or equal to 1 ms could not be reliably discriminated). This outcome also suggests that the notched noise degrades the processing of ITDs similarly for SAM and transposed tones. At this time, we can offer no explanation for Dreyer and Oxenham's listeners' inability to process ITDs conveyed by SAM tones in the presence of notched noise while being able to do so for the transposed tones.

Another difference between the two studies concerns potential explanations for why the presence of the notched noise increases threshold ITDs for both SAM and transposed tones. Dreyer and Oxenham (2008) interpreted those increases as resulting solely from a loss of off-frequency ITD information conveyed by spectral regions of the cochlear partition tuned either above or below the on-frequency 4-kHz region. Based on the notion that upward spread of excitation is greater than is downward spread of excitation, we suggest instead, or at least in addition, that increases in threshold ITD produced by notched noise may reflect an upward spread of energy from the low-pass segment of the noise into, and perhaps above, the 4-kHz target region.

At this time, we know of no psychophysical evidence to counter the parsimonious view that off-frequency listening, when it does occur for the discrimination of ITDs within high-frequency SAM and transposed tones, relies on the use of ITD information present within spectral regions of the cochlear partition tuned to frequencies above the center frequency of the target. Said differently, at this time, there appears to be no compelling evidence that listeners use off-frequency ITD information conveyed by spectral regions centered below the spectral locus of the target. By this, we certainly do not mean to imply that listeners absolutely cannot or do not utilize off-frequency ITD information within regions tuned below the target region. In order to make strong conclusions, it would seem necessary to gather data concerning the relative contributions of off-frequency listening within spectral regions above or below the target frequency region in a variety of binaural tasks and with a variety of stimuli.

The levels of noise utilized in both studies were below those found by Dreyer and Oxenham (2008) that would be required to produce *monaural masking* via upward spread of the energy of the low-pass segment into the on-frequency auditory filter containing the target. The absence of monaural masking, however, does not necessarily preclude other types

of on-frequency effects (e.g., dilution of binaural cues) that could disrupt the processing of ITDs. In our view, an inability to account for any on-frequency-based degradation of the processing of ITDs via masking should not, by default, be taken as evidence that listeners, normally, in baseline conditions employ off-frequency listening via ITD information conveyed within cochlear regions tuned below the center frequency of the target.

IV. SUMMARY AND CONCLUSIONS

Threshold ITDs conveyed by the envelopes of high-frequency SAM and transposed tones were measured in baseline conditions and in the presence of spectrally flanking notched noise, the lower spectral segment of the notched noise or the upper spectral segment of the notched noise. The data indicate that (1) relative to baseline conditions, the presence of a notched noise resulted in similar relative increases of threshold ITDs for both SAM and transposed targets; (2) the presence of the high-pass segment of the notched noise resulted in greater relative increases in threshold ITDs over those obtained in baseline conditions for SAM tones as compared to transposed tones; (3) comparisons among all of the data suggest to us that both on-frequency processing and off-frequency processing of envelope-based ITDs can be disrupted by the presence of a notched noise.

ACKNOWLEDGMENTS

The authors thank three anonymous reviewers and Dr. Richard L. Freyman for their helpful comments. This research was 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.

[1]We originally became aware of the principal findings of that study when they were presented at a meeting of the Association for Research in Otolaryngology (Dreyer *et al.*, 2005). Subsequent to that meeting and well before formal publication in this journal, both laboratories exchanged relevant information concerning the details surrounding their respective procedures and, more importantly, made each other aware of their empirical results. Especially helpful to the design of the present study was Dreyer's sharing of her Master's thesis.

[2]Our choice of levels of the target (55, 75, and 85 dB SPL) was made in order to include data taken within the range of levels employed by Dreyer in her Master's thesis (30–80 dB SPL) and to include 75 dB SPL, a level often employed in our previous relevant investigations. The data formally reported by Dreyer and Oxenham (2008) and shown in Fig. 2 represent a subset (40–70 dB SPL) of the levels utilized in Dreyer's Master's thesis.

Bernstein, L. R., and Trahiotis, C. (**1994**). "Detection of interaural delay in high-frequency SAM tones, two-tone complexes, and bands of noise," J. Acoust. Soc. Am. **95**, 3561–3567.

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.

Dreyer, A. A., and Oxenham, A. J. (**2008**). "Effects of level and background noise on interaural time difference discrimination for transposed stimuli," J. Acoust. Soc. Am. **123**, EL1–EL7.

Dreyer, A. A., Oxenham, A. J., and Delgutte, B. (**2005**). "Predicting lateralization performance at high frequencies from auditory-nerve spike timing," Abstracts of the Twenty-Eighth Midwinter Research Meeting of the Association for Research in Otolaryngology, p. 976.

Egan, J. P., and Hake, H. W. (**1950**). "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. **22**, 622–630.

Keppel, G. (**1973**). *Design and Analysis: A Researchers Handbook* (Prentice-Hall, Englewood Cliffs, NJ).

Kim, D., and Molnar, C. E. (**1979**). "A population study of cochlear nerve fibers: Comparison of spatial distribution of average rate and phase locking measures of responses to single tones," J. Neurophysiol. **42**, 16–30.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Nuetzel, J. M., and Hafter, E. R. (**1976**). "Lateralization of complex waveforms: Effects of fine-structure, amplitude, and duration," J. Acoust. Soc. Am. **60**, 1339–1346.

O'Malley, H., and Feth, L. L. (**1979**). "Relationship between psychophysical tuning curves and 'suppression'," J. Acoust. Soc. Am. **66**, 1075–1087.

Schroeder, M. R., Atal, B. S., and Hall, J. L. (**1979**). "Optimizing digital speech coders by exploiting masking properties of the human ear," J. Acoust. Soc. Am. **66**, 1647–1652.

van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.

van de Par, S., Trahiotis, C., and Bernstein, L. R. (**2000**). "The use of off-frequency information in a high-frequency binaural discrimination task," J. Acoust. Soc. Am. **86**, 526–531.

Wegel, R. L., and Lane, C. E. (**1924**). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear," Phys. Rev. **23**, 266–285.

Can the binaural system extract fine-structure interaural time differences from noncorresponding frequency channels?[a)]

David A. Magezi
MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom and School of Psychology, University of Nottingham, Nottingham NG7 2RD, United Kingdom

Katrin Krumbholz[b)]
MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom

(Received 17 July 2007; revised 25 July 2008; accepted 4 August 2008)

Due to the phase differences in the basilar membrane response between neighboring places along the cochlea, it is generally assumed that the processing of interaural time differences (ITDs) in the temporal fine structure relies on comparisons between corresponding frequency channels from the two ears. This study was aimed to test whether the auditory system is capable of extracting fine-structure ITDs from noncorresponding channels. For that, the ITD discrimination threshold was measured for a 500 Hz pure tone partially masked by a lowpass masker in one ear and a highpass masker in the other. The maskers were intended to obscure the apical or basal part of the tone's excitation pattern, respectively, and thus force the listener to extract ITDs from disparate channels. While the results did not allow any definite conclusions as to whether or not ITD processing in these conditions was based on cross-channel comparisons, some aspects of the data suggest that it was. Modeling simulations showed that any cross-channel comparisons would have to be limited to a fairly narrow frequency range of little more than one auditory-filter bandwidth. However, the between-channel phase differences within even such a narrow range would be sufficient to explain ITD sensitivity in neurophysiological data. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2980522]

PACS number(s): 43.66.Pn, 43.66.Ba [RLF] Pages: 3095–3107

I. INTRODUCTION

In humans, horizontal sound localization is mainly based on the microsecond differences in sound arrival time between the two ears [interaural time differences (ITDs)], produced by the path-length differences between the ears and the sound source. ITDs are processed by a comparison of the phase-locked temporal information mediated by the left and right auditory nerves. As the basilar membrane response to sound is a traveling wave, the phase of the response changes as a function of place along the membrane. This is why most models of binaural processing make the assumption that ITDs are processed in a channel-by-channel manner and that the channels, the temporal responses of which are being compared, originate from corresponding places in the two cochleae (Colburn, 1996). Jeffress (1948), for instance, proposed that ITDs are processed by means of a delay-and-coincidence mechanism, in which the signals from corresponding left- and right-ear channels are delayed relative to each other by axonal delays and then converge onto neurons that are excited only by coincident input.

As the slope of the cochlear phase change is particularly steep around the point where the traveling wave reaches its maximum (i.e., where the auditory response is most sensitive), the spatial correspondence between converging left- and right-ear channels would need to be very precise. A mismatch between channels would add an internal phase difference to the external ITD and would thus distort the ITD estimate. Bonham and Lewis (1999) used a gammatone-filter model based on auditory-nerve data in the cat to show that the allowable degree of mismatch between channels is 0.012 octaves (about 4.2 Hz) at 500 Hz. According to Greenwood's (1990) cochlear frequency-position function, this corresponds to a cochlear distance of only about three hair cells, implying quite a remarkable degree of precision required in the synaptic innervation of binaural neurons. During development, this kind of precision would be assumed to be achieved by a process of activity-dependent pruning of an initial larger set of connections containing both matched and unmatched pairs (for review, see Friauf and Lohmann, 1999). The pruning would be assumed to be based on the response phases of the channel pairs. In the gammatone filter, the most widely used auditory-filter model (Patterson, 1994), response phase changes by more than 180° within one equivalent rectangular filter bandwidth (ERB) (Glasberg and Moore, 1990). Any local mismatch in channel alignment would require a recalibration of the internal delay mechanism (e.g., axonal delay lines) at the relevant frequency.

While cochlear phase delays constitute a problem for models like the Jeffress model, it has been suggested that these phase delays might actually have a positive role in binaural temporal processing. Schroeder (1977) proposed

[a)]Portions of this work were presented at the 29th Midwinter Meeting of the Association for Research in Otolaryngology, Baltimore, MD, February 2006.

[b)]Author to whom correspondence should be addressed. Electronic mail: katrin@ihr.mrc.ac.uk

that it might be the phase delays between noncorresponding cochlear channels, rather than axonal delays between corresponding channels, that create the internal delays for binaural coincidence neurons. This idea was tested and expanded by Shamma *et al.* (1989) in a computational implementation, referred to as the "stereausis" model. The stereausis model assumes that there are binaural connections between both corresponding and noncorresponding channels. Thus, in this case, the developmental pruning process could be much coarser than in the case of the Jeffress or related models (e.g., driven by between-channel differences in the overall activity level rather than response phase). In this case, the fine-tuning of the mechanism would be assumed to occur at a higher stage, where the binaural activity patterns are interpreted, rather than at the stage where the patterns are generated, and would not require calibration with neurally generated delays.

An argument against the stereausis mechanism is that it may be susceptible to the changes in the cochlear phase response that are known to result from changes in sound level (Anderson *et al.*, 1971; Carney and Yin, 1988; Nuttall and Dolan, 1993; Ruggero *et al.*, 1997) and that would also be expected to occur as a result of cochlear damage. Moreover, the stereausis model would not be readily applicable to ITDs in the temporal envelope of high-frequency sounds because the temporal envelope does not exhibit the same phase differences across channels as the temporal fine structure does (e.g., Carlyon and Shamma, 2003). Thus, envelope ITDs would have to be assumed to be processed by a different mechanism than fine-structure ITDs. While this assumption is consistent with physiological results suggesting that interaural cues in low- and high-frequency sounds are processed in different structures (Joris and Yin, 1995; Batra *et al.*, 1997; for review, see Tollin, 2003), it contravenes psychophysical findings showing that envelope ITDs are processed with a similar accuracy as fine-structure ITDs (van de Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2002, 2003).

In favor of the stereausis model, Joris *et al.* (2004, 2006) pointed out that the presence of internal delays outside the physiological range (i.e., the range of ITDs encountered for a given head size; McAlpine *et al.*, 1996, 2001; Fitzpatrick *et al.*, 2000; Brand *et al.*, 2002) could be better explained by a stereausis-type model than by a Jeffress-type model. Joris *et al.* (2004, 2006) also showed that, in the cat, the dependence of the best ITD of binaural neurons in the inferior colliculus on the neurons' characteristic frequency (CF) (McAlpine *et al.*, 1996) is similar to the dependence on CF of the phase differences between responses from noncorresponding auditory-nerve fibers. However, Brand *et al.* (2002) proposed that internal delays beyond the physiological range could also be created by neural mechanisms rather than cochlear disparities. Their results suggest that internal delays are generated by contralateral inhibitory input to binaural neurons in the medial superior olive, which slightly precedes the excitatory input from the same side. Whether inhibition could explain the sometimes very large best ITDs of low-CF binaural neurons is currently a matter of debate (Joris and Yin, 2007).

A further possibility is that cross-channel comparisons may be part of a mechanism based on neurally generated internal delays, such as the Jeffress mechanism, to make it more robust against channel mismatches. For instance, in a Jeffress-type mechanism, the overall ITD could be derived by an activity-weighted average of estimates from both matched and unmatched channel comparisons. It could also be the case that both neurally generated and cochlear phase delays contribute to the internal delays for ITD processing. This idea is supported by computational analyses, which suggest that both kinds of delay may be needed to explain the ITD sensitivity of binaural neurons in mammals (Bonham and Lewis, 1999; Zhou *et al.*, 2005).

If internal delays for ITD processing are produced by stereausis, the auditory system would be expected to be able to extract ITDs from disparate channels, which can be tested by making information from corresponding channels unavailable. This approach has been used in several psychoacoustical studies with varied results. Some studies suggest that ITDs can only be extracted from corresponding channels (Toole and Sayers, 1965); other studies seem to show that ITDs can also be extracted from noncorresponding channels (Schubert and Elpern, 1959; Zerlin, 1969), but only when the frequency separation of the channels is less than one critical band (Scharf *et al.*, 1976), and yet other studies suggest that ITDs can be extracted from channels even with fairly large frequency separations (Deatherage, 1961, 1966). The main problem with all of these studies is that they used ITDs in the temporal envelopes of higher-frequency sounds, to which a stereausis-type mechanism would not be readily applicable.

The aim of the current study was to address the question of whether the auditory system is able to extract fine-structure ITDs from noncorresponding cochlear channels. For that, we measured the ITD discrimination threshold for 500 Hz pure tones, which were partially masked by a lowpass-filtered noise in one ear and a highpass-filtered noise in the other ear. The assumption was that the highpass noise would mask the basal part of the cochlear response to the tone, while the lowpass noise would mask the apical part of the response, thus forcing the listener to extract ITDs from disparate channels. The ITD discrimination threshold for these "dichotically masked" tones was compared to the threshold for "diotically masked" tones, i.e., tones that were masked by the same type of noise (lowpass or highpass) in both ears. We expected models based on a comparison of corresponding channels, such as the Jeffress model, to predict that ITD discrimination for the dichotically masked tones would be more difficult than for the diotically masked tones, which the data showed to indeed be the case; in contrast, models based on cross-channel comparisons, such as the stereausis model, were expected to predict ITD discrimination performance for the dichotic masking conditions to be similar to that for the diotic conditions. While initial modeling confirmed this expectation, and thus favored the Jeffress over the stereausis model, further simulations revealed that a more physiological version of the stereausis model was also able to predict the observed threshold difference between the diotic and the dichotic conditions. Moreover, this modified

stereausis model was able to account for individual aspects of the data, which the Jeffress model was unable to predict.

II. METHODS

A. Stimuli

ITD discrimination thresholds were measured for 500 Hz pure tones, partially masked by two independent (uncorrelated) continuous noise maskers, each presented to one ear. The experiment comprised three "diotic" masking conditions, in which the spectral composition of the maskers was the same in the two ears, and two "dichotic" conditions, in which the spectral composition of the maskers differed between the ears. The diotic conditions comprised a "lowpass," a "highpass," and an "allpass" condition. In the lowpass and highpass conditions, the maskers in both ears were lowpass or highpass filtered, respectively, with the same cutoff frequency. In the allpass condition, both maskers were neither lowpass nor highpass filtered. In the dichotic conditions, the masker was lowpass filtered in one ear and highpass filtered in the other ear; the lowpass masker was presented to either the left ("dichotic left") or the right ear ("dichotic right"). All maskers were filtered to produce a roughly constant level of excitation within their passbands. The stimuli were generated digitally at a sampling rate of 25 kHz using TDT SYSTEM 3 (Tucker-Davies Technology, Alachua, FL) and MATLAB (The Mathworks, Natick, MA). Filtering of the noise maskers was carried out in the spectral domain using 2^{18} point fast Fourier transforms. The lowpass and highpass filters were implemented as brick wall filters (i.e., setting all spectral components above the cutoff frequency to zero).

In order to maximize the chances of detecting any effect of masking condition, the ITD discrimination threshold was measured for a range of four different sensation levels (6, 8, 12, and 16 dB SL), which were the same for all masking conditions. Any effect of masking condition would be expected to disappear towards higher sensation levels, because all parts of the excitation pattern of the tone would eventually become audible at higher levels, irrespective of the spectral composition of the masker. Conversely, at very low sensation levels, any effect of masking condition might be masked by floor effects (i.e., the general difficulty of performing the task at low levels). The sensation level of the tone could be varied either by varying the level of the tone or the masker or, in the case of the lowpass and highpass maskers, by changing the cutoff frequency of the filter. As level changes can alter the transfer characteristics of the cochlear filters, sensation level was varied by adjusting the masker cutoff frequency in the case of the lowpass and highpass maskers and by changing the masker level in the case of the allpass masker. For that, the detection threshold of the tone was first determined for all diotic masking conditions (lowpass, highpass, and allpass) and in quiet. In the lowpass and highpass conditions, detection threshold was measured as a function of filter cutoff frequency. Cutoff frequencies were smaller than or equal to 500 Hz in the lowpass condition and larger than or equal to 500 Hz in the highpass condition. In order to sample the relevant parts of the masking patterns, four different cutoff frequencies were used in both the lowpass and highpass conditions, with distances of 0, 0.5, 1, and 1.5 ERBs from 500 Hz (Fig. 1; the corresponding frequency differences in hertz are shown on the top axis). In the detection threshold measurements, all maskers had the same level of about 55 dB sound pressure level (SPL) per ERB within their passbands.

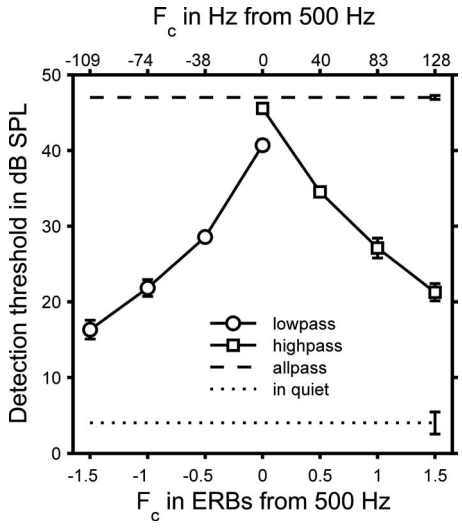


FIG. 1. Average detection thresholds for a 500 Hz pure tone in dB SPL plotted as a function of the frequency separation between the masker cutoff frequency, F_c, and the signal frequency (500 Hz) in ERBs (bottom axis) or hertz (top axis; positive and negative values represent masker cutoff frequencies above and below 500 Hz, respectively). The circles and squares show the thresholds for the lowpass and highpass masking conditions, respectively. The dashed horizontal line shows the threshold for the allpass masking condition, and the dotted line shows the threshold in quiet. Error bars show the standard error of the mean.

For the ITD discrimination threshold measurements, the sound pressure level of the tone was set to 6 dB above the detection threshold for the lowpass masker with the 500 Hz cutoff (0 ERB), yielding an average tone level of 47.4 dB SPL (std. dev.=2.3 dB). The cutoff frequencies of the lowpass and highpass maskers were then set to yield the desired tone sensation levels (6, 8, 12, and 16 dB SL) by linearly interpolating the detection threshold function for the respective masker type (Fig. 1). The resulting cutoff frequencies were used in the diotic lowpass and highpass conditions and in the dichotic conditions. The level of the lowpass and highpass maskers was the same as in the detection threshold measurements (55 dB per ERB). In the allpass condition, the desired sensation levels were achieved by reducing the masker level by the appropriate amount (based on the presentation level of the tone and its detection threshold in the 55 dB SPL allpass masker). The tone duration was 500 ms including 10 ms squared-cosine on- and off-ramps.

Stimuli were digital to analog converted with a 24 bit resolution (TDT RP2.1), amplified (TDT HB7), and presented over headphones (K240 DF, AKG, Vienna, Austria) to the participant, who was seated in a double-walled sound-attenuating room.

B. Procedure

Both the detection and ITD discrimination threshold measurements used an adaptive two-interval, two-alternative

forced-choice (2AFC) procedure with a three-down, one-up rule, which tracks 79% correct performance (Levitt, 1971). The two observation intervals in each trial had a duration of 500 ms and were separated by a gap of 500 ms. Visual feedback was provided at the end of each trial.

In the detection threshold measurements, only one of the two intervals contained a tone; any masking noise was presented continuously throughout the entire threshold run. The task was to indicate which of the two intervals contained the tone by pressing one of two response buttons. The level of the tone was changed in steps of 5 dB up to the first reversal in level, 3 dB up to the second reversal, and 2 dB for the rest of the 12 reversals that made up each threshold run. Each threshold estimate was taken as the average of the levels at the last ten reversals. At least three such threshold estimates were averaged to obtain the final threshold for each condition. The order in which different conditions were tested was randomized for each of the three threshold runs.

In the ITD discrimination threshold measurements, both intervals contained a tone. The tones were gated on and off synchronously at the two ears, but their fine-structure contained an ITD. The ITD was of the same magnitude in both intervals and was leading at the right ear in one interval and at the left ear in the other interval. The task was to indicate which of the two intervals contained the rightmost sound. The ITD between the two tones, ΔITD, was reduced and increased by a factor, v. Up to the first reversal, v was equal to 2; it was reduced to 1.5 up to the second reversal and set to 1.3 for the following ten reversals. Each threshold estimate was taken as the geometric mean of the ITD, ΔITD, for the last ten reversals. The ITD discrimination threshold for each condition was the mean of at least three such threshold estimates. As for the detection thresholds, the order in which conditions were tested was randomized. If, during the adaptive track, ΔITD exceeded 200 μs (one-tenth of the period of the 500 Hz signal) on three consecutive trials, the run was terminated and no threshold estimate was recorded for that run.

C. Data analysis

The individual detection thresholds were submitted to a two-way repeated-measures analysis of variance (ANOVA) with masking condition (lowpass and highpass) and distance of masker cutoff frequency from the 500 Hz signal frequency (0, 0.5, 1, and 1.5 ERBs, see Sec. II A) as independent within-participant factors. To test for individual- and group-level effects in the ITD discrimination threshold data, individual and average threshold estimates of individual participants were submitted to two-way repeated-measures ANOVAs with masking condition (lowpass, highpass, allpass, dichotic left, and right) and tone sensation level (6, 8, 12, and 16 dB SL) as independent within-participant factors.

D. Participants

A total of seven participants was initially recruited, five of whom (one male and four female, aged between 24 and 44 years) completed the study. All participants had absolute thresholds of 25 dB hearing level (HL) or less at audiometric frequencies and had no history of hearing or neurological disorders. One of the two participants who did not complete the study became unavailable, and the other remained unable to reliably attain a threshold estimate in one of the most difficult conditions (dichotic left at 6 dB SL) even after extensive training. Two of the remaining five participants were the authors; the other three were paid for their services at an hourly rate. Two of the five participants who completed the study were experienced in ITD discrimination and could do the task without any prior practice. The other participants underwent between 4 and 15 h of training, starting with the easiest conditions (the diotic conditions at 16 dB SL) and gradually moving on to the more difficult ones. The experimental procedures were approved by the Ethics Committee of the Nottingham University School of Psychology.

E. Additional ITD matching experiment

During the ITD discrimination task with the dichotic maskers, participants reported perceiving the tone as being lateralized towards the ear that received the lowpass masker (see Sec. III A). In order to verify these subjective reports more formally and quantify the degree of lateralization produced by the dichotic maskers, an additional experiment was conducted to measure the external ITD necessary to bring the dichotically masked tones to midline. This "matching ITD" was determined with an adaptive one-interval, two-alternative forced-choice procedure. On each trial, a single 500 ms tone was presented and the task was to indicate whether the tone was perceived to the left or to the right of the midline. The tone contained an ITD, which was changed adaptively according to the participants' responses. As this was a subjective task, no feedback was provided.

Each matching run consisted of two adaptive tracks, which were randomly interleaved. One of the tracks, referred to as the "down track," used a two-down one-up rule (the ITD of the tone was decreased after two consecutive "right" responses and increased after each "left" response) to estimate the ITD that would yield 70.7% right responses. The other track, referred to as the "up track," used a two-up one-down rule (the ITD was increased after two consecutive left responses and decreased after each right response) to estimate the ITD yielding 70.7% left responses. The starting ITD was +500 μs for the down track and −500 μs for the up track. In both tracks, the step size of the ITD increments and decrements was 250 μs up to the first reversal in ITD, 125 μs up to the second reversal, and 62.5 μs for the rest of the eight reversals that made up each track. The ITDs at the last six reversals of both tracks were averaged to yield an estimate of the matching ITD. At least three such estimates were averaged to obtain the matching ITD for each condition. The order in which different conditions were tested was randomized. If, during either of the adaptive tracks, the magnitude of the ITD exceeded 1000 μs on three consecutive trials, the run was terminated and no matching ITD estimate was recorded for that run.

At the time when the ITD matching experiment was conducted, only three of the five participants who had taken part in the main experiment were still available for testing

(the two authors and one of the other participants). For these three participants, the masking conditions tested in the ITD matching experiment were identical to the dichotic masking conditions used in the ITD discrimination threshold measurements of the main experiment (lowpass masker presented either to the left or to the right ear; masker cutoff frequencies set to yield four different tone sensation levels of 6, 8, 12, and 16 dB). In order to verify the results of these three participants, five new participants (three male, two female, aged between 23 and 34 years, with normal hearing at audiometric frequencies), four of whom were experienced in psychoacoustical tasks, were recruited to this experiment. For these new participants, the matching ITD was measured for only one tone sensation level of approximately 10 dB. In this case, the masker cutoff frequencies were not based on individual detection threshold data but were derived from the average detection threshold function measured in the main experiment. The procedure to derive the cutoff frequencies was the same as in the main experiment. The cutoff frequencies of the lowpass and highpass maskers were separated from the signal frequency of 500 Hz by 0 and 0.219 ERBs (0 and 17.4 Hz), respectively, and the tone was presented at a level of 50.7 dB. Neither the three original nor the five new participants needed any training to perform the ITD matching task. General aspects of the stimulus presentation were the same as in the ITD discrimination measurements of the main experiment.

The matching ITDs of the three participants who had also taken part in the main experiment were submitted to a two-way repeated-measures ANOVA with masking condition (dichotic left and right) and tone sensation level (6, 8, 12, and 16 dB) as independent within-participant factors. The matching ITDs of the five new participants were submitted to a paired *t*-test.

III. RESULTS AND INTERIM DISCUSSION

A. Group results

The detection thresholds were very consistent across participants, and so, only the average thresholds are shown in Fig. 1. In Fig. 1, the average thresholds for the lowpass and highpass conditions are plotted as a function of the difference between the masker cutoff frequency, F_c, and the signal frequency, F_s (500 Hz), in ERBs. When the masker cutoff frequency was equal to the signal frequency ($F_c - F_s = 0$ ERB), the detection threshold for the highpass condition (squares in Fig. 1) was similar to that of the allpass condition (dashed horizontal line), and the threshold for the lowpass condition (circles) was on average 4.9 dB lower than for the highpass condition. This asymmetry between the lowpass and highpass conditions may be related to the reported asymmetry in psychophysical suppression (Houtgast, 1972, 1973; Shannon, 1976), wherein higher frequencies are more effective at suppressing lower frequencies than vice versa. The asymmetry may also be due to the presence of nonlinear distortion products at and around the signal frequency, which would be expected to be present in the highpass but not the lowpass condition (Plomp, 1965; Greenwood, 1971; Wiegrebe and Patterson, 1999). While the phenomenon of the upward spread of masking would predict a threshold asymmetry in the opposite direction, the masker level used in the current experiment would be deemed too low to elicit this effect (Egan and Hake, 1950). As expected, the detection threshold in the lowpass and highpass conditions decreased with increasing separation between the signal and the filter cutoff frequency [$F(3,12)=2362.471$, $p<0.001$]. The detection threshold for the allpass condition (dashed horizontal line) was an average of 43 dB higher than the detection threshold in quiet (dotted horizontal line).

The ITD discrimination threshold data revealed significant main effects of both sensation level [$F(3,12)=129.887$, $p<0.001$] and masking condition [$F(4,16)=20.788$, $p<0.001$] in the group-level analysis (see Sec. II C; Fig. 2). According to Fisher's probable least significant difference (PLSD) *post hoc* tests, the main effect of masking condition was mainly due to the thresholds for the dichotic conditions (dichotic left and right) being larger than those for the diotic conditions (lowpass, highpass, and allpass; compare filled and open symbols in Fig. 2; $p<0.001$, for all diotic-dichotic comparisons, except for the comparison between dichotic left and highpass, where $p=0.006$). The main effect of sensation level was due to a substantial improvement in ITD discrimination performance with increasing sensation level, as would be expected.

Importantly, in the dichotic masking conditions, participants reported perceiving the tone as being lateralized towards the ear that received the lowpass masker, whereas the tone was perceived centrally in the diotic conditions. This pattern of lateralization would be expected if participants were extracting ITDs from disparate channels in the dichotic conditions because the lowpass masker leaves audible the basal part of the tone's excitation pattern, the phase of which leads that of the apical part. These subjective reports were confirmed and quantified by an additional ITD matching experiment, which measured the external ITD that would be necessary to compensate for the lateralization produced by the dichotic maskers (see Sec. II E). Figure 3 shows that the matching ITD strongly depended on the side to which the lowpass and highpass maskers were presented [filled symbols: original participants, $F(1,2)=59.514$, $p=0.016$); open symbols in panel (d): new participants, $t(4)=5.238$, $p=0.06$]. When the lowpass masker was presented to the left ear (dichotic-left condition, left-pointing triangles in Fig. 3), a large positive ITD (237 μs) was necessary to bring the dichotically masked tones to midline. The opposite was true for the dichotic-right condition; in this case, a large negative ITD (−398 μs) was necessary to center the masked tones (right-pointing triangles in Fig. 3). These results confirm participants' reports that, without an external ITD, the tones in the dichotic masking condition were perceived as being lateralized towards receiving the lowpass masker. There was a tendency, albeit nonsignificant, for the magnitude of the matching ITD to decrease towards higher sensation levels [$F(3,6)=1.614$; $p=0.282$].

It is unlikely that the lateralization of the signal in the dichotic conditions explains the large ITD discrimination thresholds in these conditions because the ITD discrimination threshold for pure tones is largely independent of ITD as

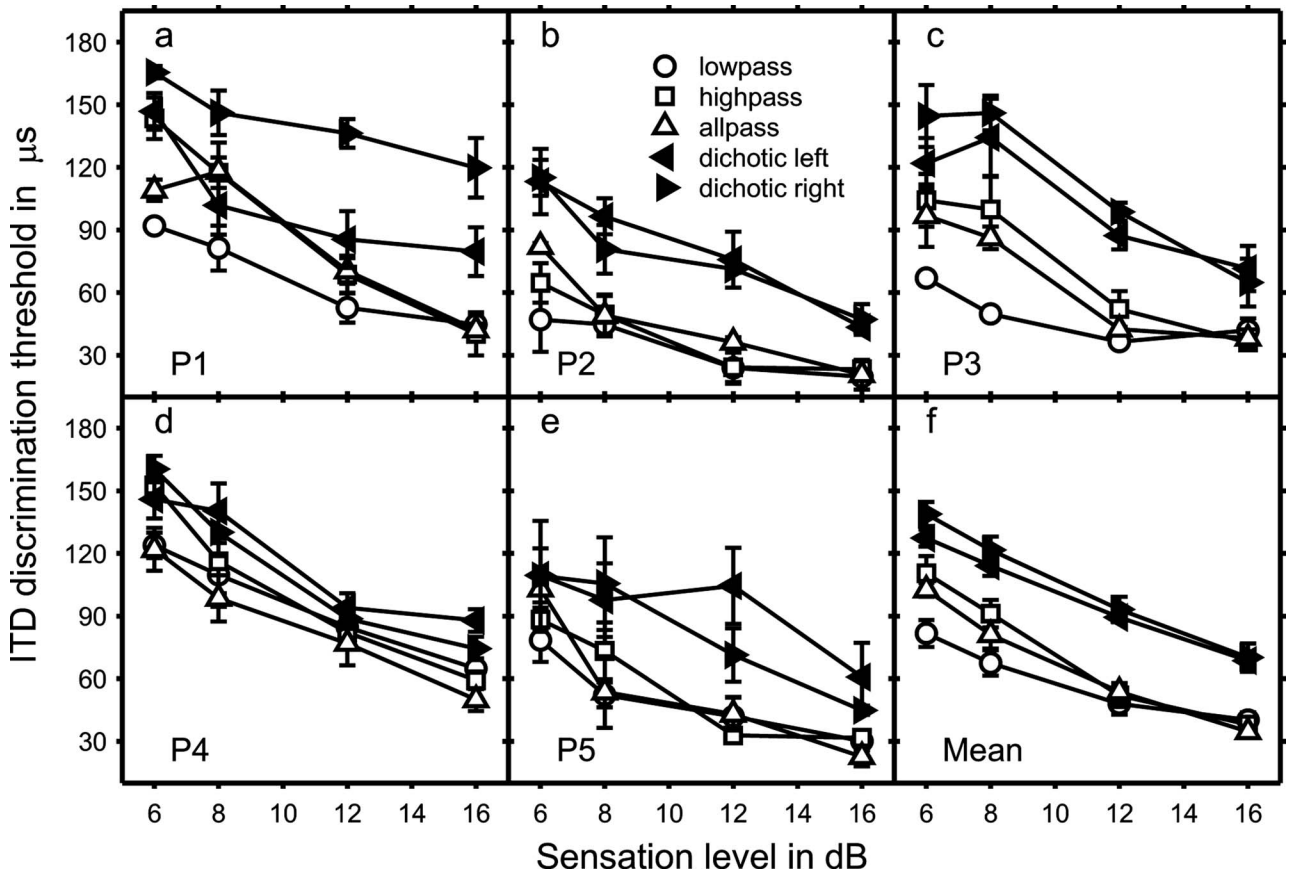


FIG. 2. Individual [panels (a)–(e)] and average [panel (f)] ITD discrimination thresholds for a partially masked 500 Hz pure tone in microseconds plotted as a function of the tone's sensation level in decibels. The parameter is the masking condition: the diotic conditions are shown by the open symbols (lowpass: circles; highpass: squares; allpass: upward-pointing triangles), and the dichotic conditions are shown by the filled triangles [dichotic left: left-pointing triangles; dichotic right: right-pointing triangles; see legend in panel (b)]. Error bars show standard errors.

long as the ITD is not ambiguous (Domnitz and Colburn, 1977; see, however, Mossop and Culling, 1998).

At first sight, the finding of larger ITD discrimination thresholds in the dichotic compared to the diotic conditions would suggest that the auditory system extracts ITDs from corresponding channels only. However, Sec. IV shows that the data can equally be accounted for by an ITD processing model based on cross-channel comparisons if the range of cross-channel comparisons is restricted to produce a physiologically plausible range of internal delays.

The *post hoc* tests showed a marginally significant difference between the lowpass and highpass conditions with the lowpass condition (open circles in Fig. 2) yielding smaller thresholds than the highpass condition (open squares; $p=0.048$) at the lower two of the four sensation levels. The difference between the lowpass and highpass conditions in the group data was largely due to two of the five participants [P1 and P3, panels (a) and (c) in Fig. 2], who showed substantially smaller thresholds in the lowpass condition than in the other two diotic conditions (highpass and allpass). This difference might be related to the fact that the apical part of the basilar membrane traveling-wave response to pure tones has a steeper phase gradient than the basal part of the response (Ren, 2002; van der Heijden and Joris, 2006; for review, see Robles and Ruggero, 2001). A steeper phase gradient would be expected to be detrimental to an ITD process-

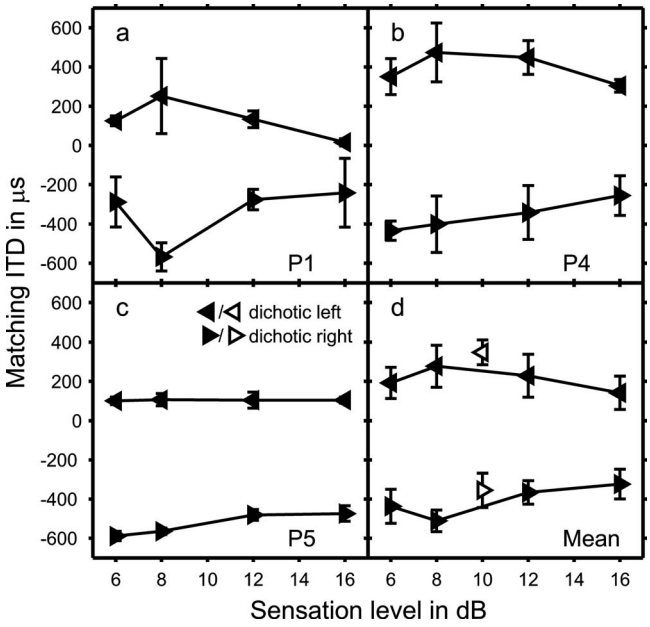


FIG. 3. External ITD, in microseconds, necessary for a 500 Hz pure tone in dichotic masking conditions to be perceived at the midline ("matching ITD"). The matching ITD is plotted as a function of the tone's sensation level in decibels. The dichotic-left and dichotic-right conditions are shown by the left- and right-pointing triangles, respectively. The filled symbols in panels (a)–(c) show individual data of three participants who had also taken part in the main experiment. The filled symbols in panel (d) show the average of these data. The open symbols in panel (d) show the data of five new participants. Error bars show standard errors.

ing mechanism that relies on the comparison between corresponding frequency channels from the two ears because mismatches in channel alignment, due to imprecisions in synaptic innervation of binaural neurons, would lead to larger errors in the ITD estimate, the larger the phase gradient of the cochlear response. In contrast, a steep phase gradient would be beneficial for a cross-channel ITD processing mechanism because the phase gradient is what produces the internal delays in a cross-channel mechanism. Thus, if the small difference between the diotic lowpass and highpass conditions observed in the current data is due to the difference in phase gradient between the basal and apical parts of the cochlear traveling-wave response, this difference would argue in favor of a corresponding-channel ITD mechanism. However, the difference could also be due to differences in the shape of the residual excitation patterns of the partially masked tones between these masking conditions. Model simulations presented in Sec. IV A 2 showed that the residual excitation patterns were broadest in the lowpass condition (see Fig. 6), which may explain why this condition yielded the smallest ITD discrimination thresholds.

B. Individual results

While participant 4's individual data showed a significant main effect of masking condition [$F(4,12)=4.454$, $p=0.02$] and the general pattern of this participant's results was consistent with that seen in the average results, participant 4 showed a much smaller threshold difference between the diotic and dichotic conditions than the other participants [Fig. 2(d)]. Participant 4 also had the highest audiometric thresholds at 500 Hz (about 25 dB HL). However, it is unlikely that these elevated audiometric thresholds would account for the anomalous ITD discrimination results because participant 4's masked thresholds were very similar to the average masked thresholds. The fact that the ITD discrimination threshold of participant 4 showed a similar decrease with increasing sensation level as the average threshold [compare Fig. 2(f)] suggests that the anomalous pattern of results was not simply due to a floor effect.

Participant 1, surprisingly, showed a substantial asymmetry between the two dichotic conditions (dichotic left and right; $p<0.001$) in that ITD discrimination performance in the dichotic-right condition [lowpass masker in the right ear; see right-pointing filled triangles in Fig. 2(a)] was much worse than in the dichotic-left condition (lowpass masker in left ear; left-pointing triangles); performance in the dichotic-left condition was almost as good as in the diotic highpass and allpass conditions (open squares and upward-pointing triangles). Such an asymmetry would be difficult to explain in an ITD model based on comparisons between corresponding channels only. In contrast, in Sec. IV C, we show that any asymmetries between the dichotic-left and dichotic-right conditions can be readily accounted for by a stereausis-type mechanism by making assumptions about the frequency distribution of binaural connections between disparate channels.

Participant 1 had required a considerable amount of training to reliably attain a threshold estimate in the dichotic-right condition at 6 and 8 dB SL. This prompted us to investigate whether the excluded participant had shown a similar asymmetry as participant 1 and whether the asymmetry had contributed to the exclusion. For that, we calculated the ratio of failed (terminated; see Sec. II B) to successful threshold runs for the dichotic-left and dichotic-right conditions at the lower two sensation levels (6 and 8 dB SL). The difference between these failure-to-success ratios (FSRs) for the dichotic-right and dichotic-left conditions (right-left FSR) is shown in Fig. 4 (open bars, right ordinate). The figure also shows the difference between the ITD discrimination thresholds for the right and left dichotic conditions as a percentage of their mean (averaged across the lower two sensation levels) for the five participants who completed the experiment (participants 1–5; filled bars, left ordinate). The figure shows that the excluded participant (P6) indeed showed a large asymmetry in the FSR between the left and right dichotic conditions; this participant successfully completed only 6 of a total of 48 threshold runs in the dichotic-left condition compared to 13 out of 18 successful runs in the dichotic-right condition. Note that the excluded participant's asymmetry was even larger than and opposite to that observed in participant 1 and that the asymmetry was effectively the reason for exclusion. The presence of asymmetry did not seem to be related to the degree of handedness; according to the Edinburgh inventory (Oldfield, 1971), participants 1 and 6, who showed an asymmetry between the dichotic conditions, exhibited the highest (100) and lowest (66) laterality ratios of the group, respectively.

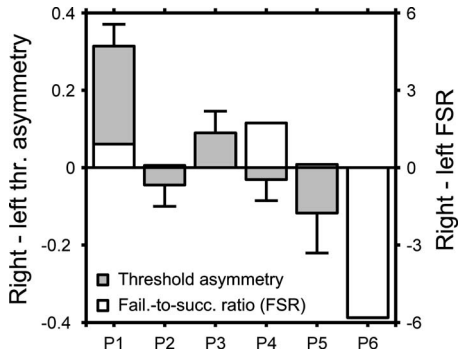


FIG. 4. Asymmetry in ITD discrimination performance between dichotic-right and dichotic-left conditions. The gray bars show the difference in ITD discrimination threshold between the dichotic-right and dichotic-left conditions expressed as a proportion of their mean and averaged across the lower two sensation levels (left ordinate) for each of the five participants who completed the ITD discrimination experiment (P1–P5). The open bars show the difference FSR (see text) between the dichotic-right and dichotic-left conditions (right ordinate) for each of six participants, one of whom (P6) did not complete the experiment (see text).

IV. MODEL SIMULATIONS

Section III showed that the ITD discrimination threshold for a partially masked 500 Hz pure tone was on average more than 50% larger for the spectrally dichotic (dichotic left and right) than for the diotic maskers (lowpass, highpass, and allpass). The difference between the dichotic and diotic masking conditions was largely independent of the tone sensation level over the range of sensation levels tested (6–16 dB SL). In this section, the measured ITD discrimination

thresholds are compared to predictions from two opposing models of ITD processing, one representing ITD processing based on interaural temporal comparisons between corresponding frequency channels from the two ears by means of neurally generated internal delays and the other representing processing based on comparisons between noncorresponding channels with internal delays being generated mechanically at the level of the cochlea (see Sec. I). The first model was a computational version of Jeffress' (1948) delay-and-coincidence mechanism, in which the cross-correlation function between corresponding frequency channels was computed for a range of correlation lags, assumed to be created by axonal propagation delays (APDs). This model will henceforth be referred to as the APD model. The second model was a computational version of Schroeder's (1977) cross-channel mechanism, in which the cross-correlation at lag zero was calculated between all possible channel combinations within a range of frequencies around the signal frequency (500 Hz). This model will henceforth be referred to as the cross-channel (XCH) model.

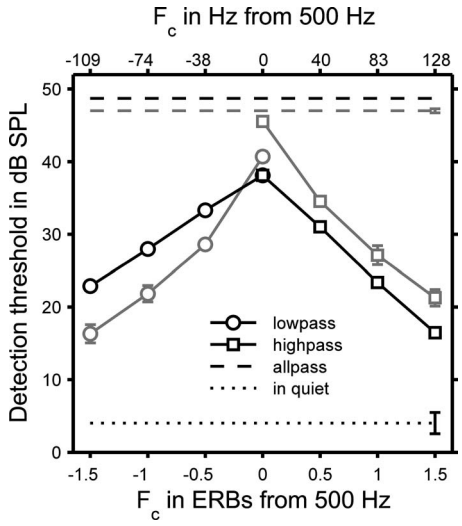


FIG. 5. Model simulations of the detection threshold data from Fig. 1. The simulated thresholds are shown by the black symbols and lines and plotted in the same way as the measured thresholds in Fig. 1. The measured thresholds are replotted for comparison (gray symbols and lines). The parameter is the masking condition (see legend).

A. Model architecture

Both models (APD and XCH) consisted of four stages: the first and second stages simulated the peripheral response to the pure-tone signal and the effect of the noise masker, the third stage simulated the binaural processing of the signals from the left and right ears, and the last stage calculated a decision measure upon which the threshold estimates were based.

1. Peripheral response to pure-tone signal

The peripheral response to the pure-tone signal was modeled using the DSAM/AMS software package (O'Mard and Meddis, 2004). The spectral analysis performed by the cochlea was simulated by a 65 channel gammatone filterbank with center frequencies between 200 Hz and 1 kHz (±5 ERBs around 500 Hz), evenly distributed on an ERB scale (6.5 channels per ERB). The filterbank output was half-wave rectified and fourth order lowpass filtered at 1 kHz to simulate the transformation from the mechanical response of the basilar membrane to the neural activity pattern (NAP) flowing up the auditory nerve. The first 50 ms of the NAP were discarded to allow the model response to reach a steady state. Finally, each channel output was normalized by its root-mean-square (rms) amplitude.

2. Effect of noise masker

The effect of the noise masker was modeled by first simulating the detection thresholds using Glasberg and Moore's (1990) excitation-pattern model. The detection thresholds were simulated by calculating the positive difference, $R=\max(NS-N,0)$, between the excitation pattern for the tone signal plus noise (NS; expressed in dB per ERB) and for the noise alone (N); R is referred to as the residual excitation pattern. The noise, N, was the sum of the external noise and an internal noise, N_0, which had a constant level per ERB. The average tone detection threshold in quiet (dotted horizontal line in Fig. 1) was used to calculate the internal noise level, N_0. All excitation patterns were calculated with 512 channels between 20 Hz and 1.9 kHz (±10 ERBs around 500 Hz), evenly distributed on an ERB scale (25.6 channels per ERB). Model detection threshold was defined as the tone level at which the residual excitation pattern reached a criterion area, A_{crit}, which was a free parameter in the fitting process. All conditions of the detection threshold measurements were fitted simultaneously with the same value of A_{crit}, which was varied to minimize the rms deviation between the simulated and observed thresholds. Figure 5 shows that the model (black symbols and lines) produced a reasonably good fit to the detection threshold data (gray symbols and lines). As would be expected based on the linear nature of the model, the model was unable to simulate the difference between the lowpass and highpass conditions at masker cutoff frequencies close to the signal frequency.

The simulated detection thresholds were then used to calculate the cutoff frequencies for the lowpass and highpass maskers and the levels for the allpass masker to yield the appropriate tone sensation levels for the simulation of the ITD discrimination thresholds (6, 8, 12, and 16 dB SL); this was done in exactly the same was as in the experiment (see Sec. II A). Finally, residual tone excitation patterns were calculated for all masker types (lowpass, highpass, and allpass) and sensation levels (6, 8, 12, and 16 dB SL). The residual excitation patterns were converted to linear amplitude units normalized to the maximum of the pattern with the highest peak (i.e., the pattern for the highest sensation level in the allpass condition; see Fig. 6) and interpolated to the channel frequencies of the gammatone filterbank used in the first stage of the simulation. To account for the fact that the ITD discrimination threshold functions reach an asymptote at high sensation levels, the residual excitation patterns were compressed by exponentiation with a parameter $\alpha<1$, which was a free parameter in the ITD discrimination threshold simulation. Compression makes the excitation patterns for different sensation levels more similar. The normalization of

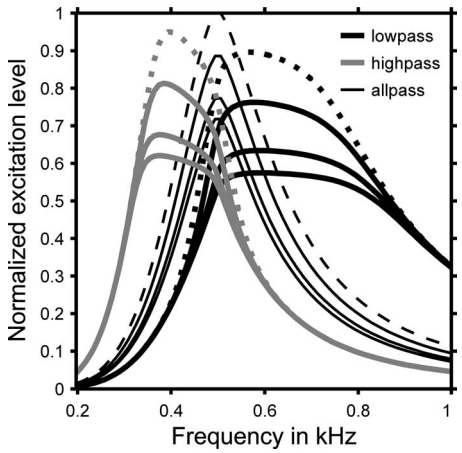


FIG. 6. Simulated residual excitation patterns of a 500 Hz pure tone in different masking conditions and for different sensation levels (bold black lines: lowpass; bold gray lines: highpass; thin black lines: allpass). Excitation level was expressed in linear units and normalized to the maximum of the pattern with the highest peak, i.e., the pattern for the highest sensation level in the allpass condition (dashed thin black line). The dotted bold black and gray lines show the patterns for the lowpass and highpass conditions, respectively, for the highest sensation level; these patterns were used to create the cross-correlation patterns shown in Figs. 7(d) and 8(d) (see text).

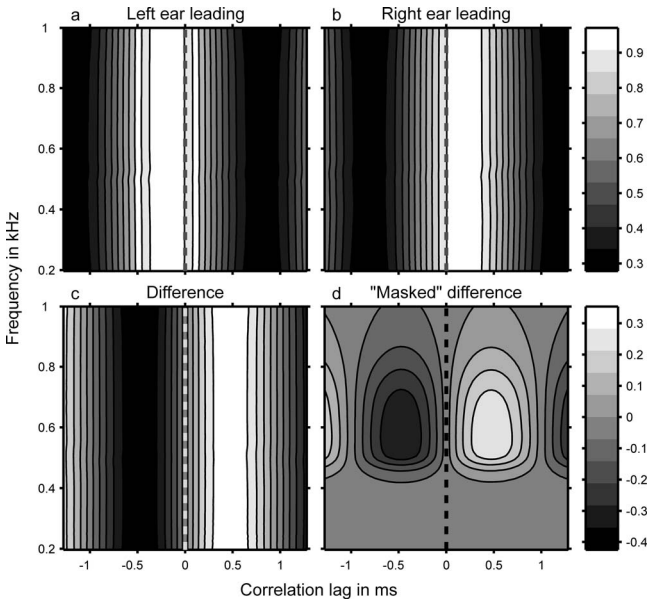


FIG. 7. APD model. Panels (a) and (b) show the cross-correlation patterns of the normalized NAPs for a 500 Hz tone leading by 190 μs at the left and right ears, respectively; different correlation values are represented by different gray shades [see color bar to the right of panel (b)]. Panel (c) shows the difference between the patterns in panels (a) and (b) [(b)-(a)] and panel (d) shows the difference pattern after weighting with the residual excitation patterns for the diotic lowpass condition at 16 dB SL (shown by bold black dotted line in Fig. 6); again different values of the difference are represented by different gray shades [see color bar to the right of panel (d)]. The dashed vertical lines mark the point at which the internal (cross-correlation) delay is zero.

the patterns to a maximum value of unity means that the compression did not change the codomain of the patterns. The patterns shown in Fig. 6 are uncompressed.

3. Binaural processing

In the next stage, the normalized NAPs from each ear were weighted by the appropriate residual excitation pattern from stage 2 and then combined to extract the interaural temporal information contained in the resulting patterns.

For the APD model, each channel from the left ear was cross-correlated with the corresponding channel from the right ear over a range of lags between ± 1280 μs by integrating the cross-product between the non-mean-corrected channel waveforms (see Bernstein and Trahiotis, 1996) over time. Panels (a) and (b) in Fig. 7 show the cross-correlation patterns of the normalized NAPs for 500 Hz tones leading by 190 μs at the left (a) or right ear (b); to show the entire pattern, the NAPs were not yet weighted by the residual excitation patterns in this example. The figure reveals a vertical ridge of high correlation, the horizontal position of which reflects the ITD of the tone [white-shaded areas in Figs. 7(a) and 7(b)].

For the XCH model, each channel in the left-ear NAP was correlated (lag zero) with all channels in the right-ear NAP, again using the cross-product between the non-mean-corrected waveforms. Panels (a) and (b) in Fig. 8 show that the ridge of high correlation along the central diagonal (marked by a dashed line) in the cross-channel correlation pattern of a 500 Hz tone is shifted towards the upper left corner of the pattern when the tone is leading in the left ear (by 190 μs as in Fig. 7) and towards the lower right corner when the tone is leading in the right ear. The central diagonal represents points where the left- and right-ear channels have the same frequencies. Panel (c) in Figs. 7 and 8 shows the difference between the cross-correlation patterns for the right-leading and left-leading tones shown in panels (b) and (a). These difference patterns simulate the information potentially available to the binaural system in the 2AFC paradigm used in the ITD discrimination threshold measurements. The effect of the masking noise would be to obscure part of these patterns. Panel (d) in Figs. 7 and 8 shows the difference patterns from panel (c) after weighting with the residual excitation patterns. The difference pattern in Fig. 7(d) was weighted with the residual excitation pattern for the diotic lowpass condition at 16 dB SL (see bold black dotted line in Fig. 6); the figure shows that this masking condition obscures the low-frequency part of the difference pattern. In Fig. 8(d), the difference pattern was weighted with the residual excitation pattern for the 16 dB SL dichotic-left condition (see bold black and gray dotted lines in Fig. 6), which obscures the low frequencies in the left ear and the high frequencies in the right ear and thus limits the difference pattern to the lower right quadrant.

4. Decision measure

For a given value of the compression exponent, α, difference patterns, as shown in Figs. 7(d) and 8(d), were calculated for all masking conditions and sensation levels, using ten different ITDs, equally spaced between 10 and 190 μs [Figs. 7(d) and 8(d) show examples for the 190 μs ITD]. The decision measure, D, was the square root of the integral of the squared difference patterns. Threshold was defined as the ITD at which D reached a criterion value, D_{crit}, which was a free parameter of the fitting process and was chosen to

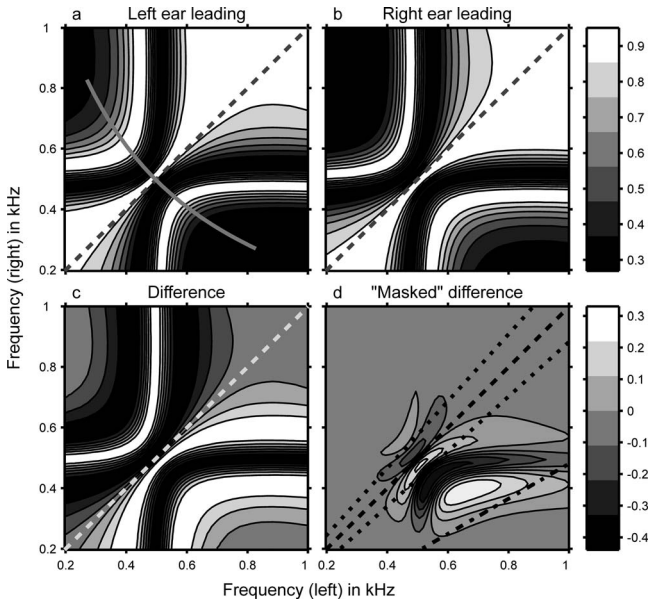


FIG. 8. Cross-channel (XCH) model plotted in the same way as the APD model in Fig. 7. Panels (a) and (b) show the cross-correlation patterns for the same left- and right-leading tones as used in Fig. 7, panel (c) shows the difference of the patterns shown in panels (a) and (b) ([(b)-(a)], as in Fig. 7), and panel (d) shows the difference pattern after weighting with the residual excitation patterns for the dichotic-left condition at 16 dB SL (see dotted bold black and gray lines in Fig. 6). The dashed diagonal lines represent points where the channel frequencies from the left and right ears are equal. The dotted lines to the left and right of the diagonal in panel (d) represent the boundaries of the frequency band within which the phase differences between the left- and right-ear channels range between $\pm\pi$ rad of their respective average frequency (in ERBs; see also Fig. 10). The dash-dotted line shows the outer boundary of one of the two SCBs, where the phase differences between the left- and right-ear channels range from π to 3π. The gray curved line in panel (a) shows points covered by the cross-section of the cross-correlation pattern shown in Fig. 10.

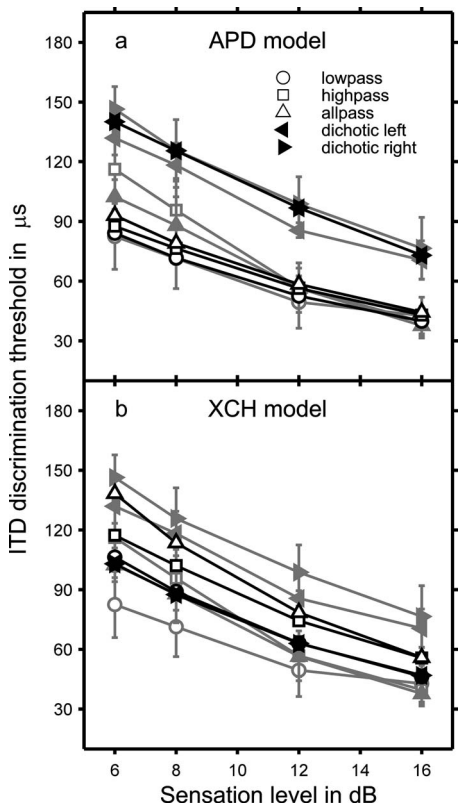


FIG. 9. Simulated ITD discrimination thresholds (black symbols and lines) based on the APD (a) and XCH (b) models plotted in the same way as the measured thresholds in Fig. 2. The parameter is the masking condition (see legend). The average measured thresholds were replotted from Fig. 2(f) for comparison (gray symbols and lines).

minimize the rms deviation between predicted and observed thresholds. The other free parameter was the compression exponent, α. The value of α was determined by repeating the fitting process for each of 31 values of α, equally spaced between 0.1 and 0.25, and choosing the value that minimized the rms deviation between predicted and observed thresholds. Smaller values of α (more compression) made the slope of the function relating the simulated ITD thresholds to the tone sensation levels shallower.

B. Simulation results and interim discussion

The black symbols and lines in Fig. 9(a) show that the APD model produced a remarkably good fit to the experimental data (gray symbols and lines); the rms deviation between the simulated and measured thresholds amounted to only 8.5 μs. The best-fitting compression exponent, α, was 0.19 in this simulation. The model yielded similar thresholds for all three diotic masking conditions (lowpass, highpass, and allpass; open symbols) and correctly predicted larger thresholds for the dichotic conditions (filled symbols). The difference between the simulated dichotic and diotic thresholds was similar to that in the data. In contrast, the XCH model provided a poor fit to the data (rms deviation =24.6 μs), predicting largely similar thresholds for all conditions [Fig. 9(b)], with the smallest thresholds for the dichotic and the lowpass conditions (filled triangles and open circles) and the largest thresholds for the diotic allpass condition (open triangles). The best-fitting α amounted to 0.22 in the XCH simulation. Note that, while the corresponding-channel model used in the current simulations was based on Jeffress' (1948) delay-and-coincidence mechanism and thus involved APDs, similar findings would also be expected to apply to other models, where ITDs are extracted from corresponding channels, but the ITD analysis is based on another mechanism, such as inhibition (Brand *et al.*, 2002; for review, see Grothe, 2003; McAlpine and Grothe, 2003; and Joris and Yin, 2007).

At first glance, these modeling results seem to provide strong evidence for a mechanism based on comparisons between corresponding frequency channels. However, a possible reason for the poor performance of the XCH model may have been that the model included more cross-channel information than necessary, which unduly benefited the dichotic conditions. The cross-channel correlation patterns of the XCH model contained a relatively high degree of informational redundancy, in that, in addition to the correlation peak near the diagonal [white shading near dashed diagonal line in Figs. 8(a) and 8(b)], the patterns contained another correlation peak (representing the first slip cycle) on either side of the diagonal (hyperbolically shaped white-shaded areas). This is illustrated in Fig. 10, which shows a slice of the cross-channel correlation pattern for a tone with zero ITD taken along the curved gray line in Fig. 8(a); this line repre-

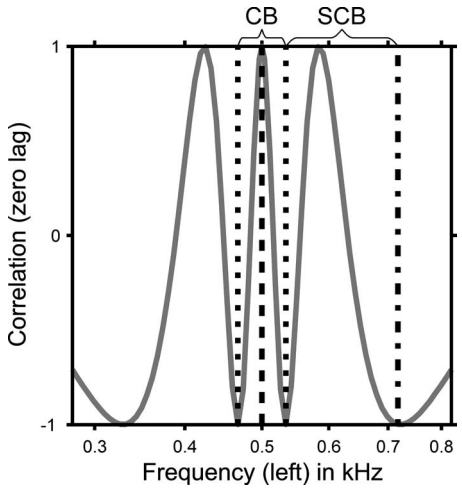


FIG. 10. Cross-section through the cross-correlation pattern for a 500 Hz tone with zero ITD taken along the curved gray line in Fig. 8(a), which represents points where the left- and right-ear channel frequencies have equal separation (in ERBs) from 500 Hz, which is where the line crosses the diagonal. The correlation values are plotted as a function of the channel frequencies in the left ear. As in Fig. 8(d), the dashed vertical line represents the point where the channel frequencies from the left and right ears are equal (500 Hz). The dotted lines show the boundaries of the frequency band within which the phase differences between the left- and right-ear channels range between $\pm\pi$ (referred to as the CB) and the dash-dotted line marks the outer boundary of one of SCBs.

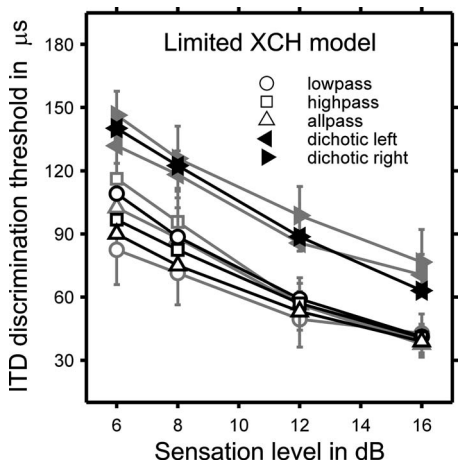


FIG. 11. Simulated ITD discrimination thresholds (black symbols and lines) based on the limited XCH model. As in Fig. 9, the gray symbols and lines show the average measured thresholds for comparison.

sents points where the left- and right-ear channel frequencies have equal separation (in ERBs) from 500 Hz, where the line crosses the diagonal.

The degree of redundancy in the XCH patterns was (largely arbitrarily) determined by the frequency range of input channels (filter frequencies in the gammatone filterbank) and may be argued to be physiologically implausible, given that the best ITDs of binaural neurons in the mammalian brainstem seem to be limited to the range of $\pm\pi$ rad of the neurons' best frequency (which excludes the slip cycles; McAlpine *et al.*, 1996). The XCH simulation was therefore repeated, this time limiting the cross-correlation patterns to a physiologically more plausible range. For that, the XCH patterns were limited to a band around the diagonal, shown by the dotted lines to the left and right of the diagonal in Fig. 8(d), and rerunning the simulation as before with the limited patterns. Within the band, the phase differences between the left- and right-ear channels ranged between $\pm\pi$ rad of their respective average frequency (in ERBs; see vertical dotted lines in Fig. 10). The band was about 1.2 ERBs wide. Figure 11 shows that the limited XCH model indeed produced a better fit to the data. Like the APD model, the limited XCH model yielded the largest thresholds for the dichotic conditions, and the difference between the simulated thresholds for the dichotic and the diotic conditions was similar to that in the data. The rms deviation between predicted and measured thresholds amounted to only 10.4 μs, which is comparable to that for the APD model. The best-fitting compression exponent, α, for the limited XCH simulation was similar to that for the original XCH simulation (0.215).

C. Accounting for the asymmetries between the dichotic conditions in individual data sets

By limiting the frequency range of cross-channel comparisons, the XCH model could be made to provide as good an account of the dichotic ITD discrimination thresholds as the APD model. In a similar way, the XCH model can also be made to explain another aspect of the data, namely, the asymmetry between the thresholds for the dichotic-left and dichotic-right conditions observed in some participants (see Sec. III B), by limiting the cross-channel correlation patterns to an asymmetrical band around the central diagonal. In particular, a band that includes the slip cycle on the right of the diagonal (higher frequencies in the left ear and lower frequencies in the right ear) would be expected to predict better performance for the dichotic-left (lowpass masker in the left ear) than the dichotic-right condition, whereas a band that includes the left slip cycle would be expected to favor the dichotic-right condition. To test this, the data of participant 1, whose thresholds for the dichotic-left condition were substantially smaller than those for the dichotic-right condition [Fig. 2(a)], were fitted by limiting the XCH patterns to a band that included the slip cycle on the right but not on the left of the diagonal [see dash-dotted line in Figs. 8(d) and 10]. The relative weight of the central band (CB) and the slip-cycle band (SCB) in Fig. 10 was a free parameter in this simulation and was adjusted to minimize the rms deviation between the simulated and participant 1's individual thresholds.

Figure 12 shows that this asymmetrically weighted XCH model provided a reasonably good account of the asymmetry between the dichotic conditions in participant 1's ITD threshold data. The relative weights of the CBs and SCBs that minimized the rms deviation between simulated and measured thresholds (17.9 μs) amounted to 70% and 30%, respectively. The compression exponent, α, amounted to 0.185 in this simulation. It should be noted that the XCH model could similarly be made to explain the individual results of participant 4, who showed only marginal differences between the ITD discrimination thresholds for the dichotic and diotic conditions [Fig. 2(d)]. Individual differences and

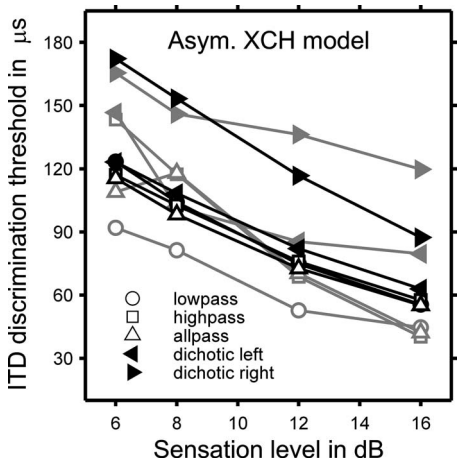


FIG. 12. Simulated ITD discrimination thresholds for participant 1 (black symbols and lines) based on the asymmetrically weighted XCH model. Participants 1's measured thresholds were replotted from Fig. 2(a) for comparison (gray symbols and lines).

asymmetries in the frequency distribution of cross-channel connections in the XCH model could arise as a consequence of incomplete pruning during development. It would seem difficult or impossible to see how any model that is purely based on comparisons between corresponding frequency channels (like the APD model) could explain the asymmetry in the effect of combining activity in different frequency channels shown by participant 1, unless comparisons between noncorresponding channels were also included in the model.

V. SUMMARY AND CONCLUSIONS

The current study showed that the ITD discrimination threshold for pure tones partially masked by highpass or lowpass noise maskers was about twice as large in the dichotic masking conditions, where the masker spectral characteristics differed between the two ears, than in the diotic conditions, where the masker had the same spectral properties in both ears. The dichotic conditions were intended to force the participants to extract interaural temporal information from disparate frequency channels in the two ears. The finding that ITD discrimination performance was poor in these conditions at first sight suggested that the auditory system extracts interaural temporal information mainly from corresponding frequency channels in the two ears. However, the model simulations showed that a computational version of Schroeder's (1977) cross-channel (XCH) model of ITD processing was able to provide a similarly accurate account of the current data as a corresponding-channel model based on Jeffress' (1948) delay-and-coincidence mechanism (APD model) if the range of cross-channel comparisons was restricted so as to produce a physiologically plausible range of internal delays. The XCH model was also able to explain the asymmetry observed in the dichotic ITD discrimination thresholds of some participants, which would have been difficult, if not impossible, to explain with the APD model, unless cross-channel comparisons were also included in the APD model.

The fact that, in the dichotic masking conditions, participants perceived the tone as being lateralized towards the ear receiving the lowpass masker suggests that participants may indeed have been using information from disparate channels in these conditions. The model results suggest that, if performance in the dichotic conditions was based on cross-channel comparisons, such comparisons would have to be limited to a fairly narrow frequency range of little more than one auditory-filter bandwidth around each channel. The modeling also showed that even within this narrow range, phase differences between noncorresponding channels cover the entire range of internal delays observed in ITD-sensitive neurons.

However, whether or not these phase delays actually contribute to the internal delays for fine-structure ITD processing cannot be determined from the current data. Conceivably, cross-channel comparisons could also be part of a mechanism based on neurally generated internal delays to make the mechanism robust against channel mismatch. Alternatively, neurally generated delays and cochlear phase delays may both contribute to ITD sensitivity at low frequencies.

Future studies could investigate the effect of sound level, which has been shown to affect cochlear phase delays, on the lateralization of partially masked tones.

ACKNOWLEDGMENTS

This work was funded by the Medical Research Council. A conference presentation of this work was supported by Deafness Research UK and The David and Frederick Barclay Foundation. We thank Richard Freyman and two anonymous reviewers for their helpful comments on previous versions of this manuscript.

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (**1971**). "Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects," J. Acoust. Soc. Am. **49**, 1131–1139.

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (**1997**). "Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex, I: Heterogeneity of responses," J. Neurophysiol. **78**, 1222–1236.

Bernstein, L. R., and Trahiotis, C. (**1996**). "On the use of the normalized correlation as an index of interaural envelope correlation," J. Acoust. Soc. Am. **100**, 1754–1763.

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.

Bernstein, L. R., and Trahiotis, C. (**2003**). "Enhancing interaural-delay-based extents of laterality at high frequencies by using "transposed stimuli"," J. Acoust. Soc. Am. **113**, 3335–3347.

Bonham, B. H., and Lewis, E. R. (**1999**). "Localization by interaural time difference (ITD): Effects of interaural frequency mismatch," J. Acoust. Soc. Am. **106**, 281–290.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature (London) **417**, 543–547.

Carlyon, R. P., and Shamma, S. (**2003**). "An account of monaural phase sensitivity," J. Acoust. Soc. Am. **114**, 333–348.

Carney, L. H., and Yin, T. C. (**1988**). "Temporal coding of resonances by low-frequency auditory nerve fibers: Single-fiber responses and a population model," J. Neurophysiol. **60**, 1653–1677.

Colburn, H. S. (**1996**). "Computational models of binaural processing," in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Poper, and R. R. Fay (Springer, New York), pp. 332–400.

Deatherage, B. H. (**1961**). "Binaural interaction of clicks of different fre-

quency content," J. Acoust. Soc. Am. **33**, 139–145.
Deatherage, B. H. (**1966**). "Examination of binaural interaction," J. Acoust. Soc. Am. **39**, 232–249.
Domnitz, R. H., and Colburn, H. S. (**1977**). "Lateral position and interaural discrimination," J. Acoust. Soc. Am. **61**, 1586–1598.
Egan, J. P., and Hake, H. W. (**1950**). "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. **22**, 622–630.
Fitzpatrick, D. C., Kuwada, S., and Batra, R. (**2000**). "Neural sensitivity to interaural time differences: Beyond the Jeffress model," J. Neurosci. **20**, 1605–1615.
Friauf, E., and Lohmann, C. (**1999**). "Development of auditory brainstem circuitry. Activity-dependent and activity-independent processes," Cell Tissue Res. **297**, 187–195.
Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Greenwood, D. D. (**1971**). "Aural combination tones and auditory masking," J. Acoust. Soc. Am. **50**, 502–543.
Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species: 29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Grothe, B. (**2003**). "New roles for synaptic inhibition in sound localization," Nat. Rev. Neurosci. **4**, 540–550.
Houtgast, T. (**1972**). "Psychophysical evidence for lateral inhibition in hearing," J. Acoust. Soc. Am. **51**, 1885–1894.
Houtgast, T. (**1973**). "Psychophysical experiments on 'tuning curves' and 'two-tone inhibition'," Acustica **29**, 168–179.
Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.
Joris, P., and Yin, T. C. (**2007**). "A matter of time: Internal delays in binaural processing," Trends Neurosci. **30**, 70–78.
Joris, P. X., Van de Sande, B., Louage, D. H., and van der Heijden, M. (**2006**). "Binaural and cochlear disparities," Proc. Natl. Acad. Sci. U.S.A. **103**, 12917–12922.
Joris, P. X., van der Heijen, M., Louage, D., Van der Sande, B., and Van Kerckhoven, C. (**2004**). "Dependence of binaural and cochlear 'best delays' on characteristic frequency," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 396–402.
Joris, P. X., and Yin, T. C. (**1995**). "Envelope coding in the lateral superior olive, I: Sensitivity to interaural time differences," J. Neurophysiol. **73**, 1043–1062.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 466–477.
McAlpine, D., and Grothe, B. (**2003**). "Sound localization and delay lines: Do mammals fit the model?," Trends Neurosci. **26**, 347–350.
McAlpine, D., Jiang, D., and Palmer, A. R. (**1996**). "Interaural delay sensitivity and the classification of low best-frequency binaural responses in the inferior colliculus of the guinea pig," Hear. Res. **97**, 136–152.
McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). "A neural code for low-frequency sound localization in mammals," Nat. Neurosci. **4**, 396–401.
Mossop, J. E., and Culling, J. F. (**1998**). "Lateralization of large interaural delays," J. Acoust. Soc. Am. **104**, 1574–1579.
Nuttall, A. L., and Dolan, D. F. (**1993**). "Two-tone suppression of inner hair cell and basilar membrane responses in the guinea pig," J. Acoust. Soc. Am. **93**, 390–400.
O'Mard L. P., and Meddis R. (**2004**). "A New visual dimension to auditory modelling using DSAM," Assoc. Res. Otolaryngol. Abstr. **27**, 312–313.
Oldfield, R. C. (**1971**). "The assessment and analysis of handedness: The Edinburgh inventory," Neuropsychologia **9**, 97–113.
Patterson, R. D. (**1994**). "The sound of a sinusoid: Spectral models," J. Acoust. Soc. Am. **96**, 1409–1418.
Plomp, R. (**1965**). "Detectability threshold for combination tones," J. Acoust. Soc. Am. **37**, 1110–1123.
Ren, T. (**2002**)."Longitudinal pattern of basilar membrane vibration in the sensitive cochlea," Proc. Natl. Acad. Sci. U.S.A. **99**, 17101–17106.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.
Scharf, B., Florentine, M., and Meiselman, C. H. (**1976**). "Critical band in auditory lateralization," Sens. Processes **1**, 109–126.
Schroeder, M. R. (**1977**). "New viewpoints in binaural interactions," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, New York), pp. 455–467.
Schubert, E. D., and Elpern, B. S. (**1959**). "Psychophysical estimate of the velocity of the traveling wave," J. Acoust. Soc. Am. **31**, 990–994.
Shamma, S. A., Shen, N. M., and Gopalaswamy, P. (**1989**). "Stereausis: Binaural processing without neural delays," J. Acoust. Soc. Am. **86**, 989–1006.
Shannon, R. V. (**1976**). "Two-tone unmasking and suppression in a forward-masking situation," J. Acoust. Soc. Am. **59**, 1460–1470.
Tollin, D. J. (**2003**). "The lateral superior olive: A functional role in sound source localization," Neuroscientist **9**, 127–143.
Toole, F. E., and Sayers, B. M. (**1965**). "Inferences of neural activity associated with binaural acoustic images," J. Acoust. Soc. Am. **38**, 769–779.
van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.
van der Heijden, M., and Joris, P. X. (**2006**). "Panoramic measurements of the apex of the cochlea," J. Neurosci. **26**, 11462–11473.
Wiegrebe, L., and Patterson, R. D. (**1999**). "Quantifying the distortion products generated by amplitude-modulated noise," J. Acoust. Soc. Am. **106**, 2709–2718.
Zerlin, S. (**1969**). "Traveling-wave velocity in the human cochlea," J. Acoust. Soc. Am. **46**, 1011–1015.
Zhou, Y., Carney, L. H., and Colburn, H. S. (**2005**). "A model for interaural time difference sensitivity in the medial superior olive: Interaction of excitatory and inhibitory synaptic inputs, channel dynamics, and cellular morphology," J. Neurosci. **25**, 3046–3058.

Binaural model for artificial spatial sound localization based on interaural time delays and movements of the interaural axis

Laurent Kneip[a)]
Friedrich-Alexander-University, Erlangen-Nürnberg, Erlangen, Bavaria 91052, Germany

Claude Baumann[b)]
Director of the Convict Episcopal, Boarding Institution, Luxembourg, L-2132

(Received 22 June 2007; revised 16 July 2008; accepted 30 July 2008)

This paper presents a mathematical model for sound localization in space using two-microphone devices that possess at least two degrees of freedom. It proves a series of theorems and lemmas that are based on time difference of arrival measurements and movements of the interaural axis, forming a powerful instrument for practical robot applications. For instance, it shows that a single determined rotation of the interaural axis is sufficient to exactly yield the azimuth or the elevation of an immobile sound source in the far field, independently of microphone spacing and the speed of sound and hence of the surrounding medium. It proves that at any moment the knowledge of one value determines the magnitude of the other, with the restriction that the sign of the second value is undefined, which means that, depending on the rotation, either the back-front or the up-down ambiguity is kept unsolved. This paper also shows that parallax motion unlocks essential information about the distance and the Cartesian coordinates of the sound source. Shifting the microphone system sideways fixes the distance and the coordinate on the interaural axis. Combining rotation and translation movements completely solves the localization problem. In order to illustrate the efficacy of the model, this paper presents experiments with a low cost robot developer kit during which the azimuth, the elevation, and the distance of continuous sound sources are determined at a precision of 10° and 0.5 m, respectively. Achieving this performance with low power material demonstrates how easily the model can be implemented into any robotic system.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977746]

PACS number(s): 43.66.Pn, 43.66.Qp, 43.60.Gk [DRD] Pages: 3108–3119

I. INTRODUCTION

In an effort to create a robot that is able to determine the direction and the distance to a sound source, this paper wants to contribute to the formal description of artificial sound localization in space. It presents a theoretical binaural model that is based on interaural time delay (ITD) cues in combination with the motion of the interaural axis.

Natural spatial hearing has largely inspired this model. The human listener's ability to localize sound in space still is best summarized in Blauert's encyclopaedia.[1] It is generally assumed that interaural level difference (ILD) and ITD cues contribute to the yielding of the horizontal angular component of the sound vector, whereas spectral head related transfer function cues help to find the vertical component and to solve the back-front ambiguity. Dynamic cues, related to head motion, are commonly accepted as complementary in clarifying remaining ambiguities. However, already van Soest[2] argued that head movements could be more useful to localize sound in space, stating that through a head movement, according to the phase theory, one receives a bundle of cones of confusion that must possess an intersectional line, which determines the direction to the sound source. Wallach,[3] another early scientist, asserted that angular displacements of the interaural axis can be effective in sound localization. He suggested that these head movements not only provide information about the azimuth but also about the elevation. Ignoring monaural cues, he supposed that head movements that do not displace the interaural axis do not contribute to the sound localization. Wallach's hypotheses have been verified and confirmed repeatedly.[4,5] Thurlow *et al.*[6] isolated three fundamental types of head rotations: *rotation* around the vertical axis, *tip* around the interaural axis, and *pivot* around the horizontal baseline, which is orthogonal to the interaural axis. They also observed various combinations of consecutive movements.

While directional localization research has quite a long history, sound distance studies are comparably recent.[7] There is some evidence that sound level cues play an important role in distance evaluation. Other monaural cues are the spectral composition[8,9] and the proportion of direct to reverberant sound intensity.[10,11] From the binaural point of view Speigle and Loomis[12] found that parallax motion of the interaural axis has weak but significant influence on distance cues. They developed a triangulation formula for sound sources in the horizontal plane that determines the distance from the change in azimuth.

Recent work only tries to imitate natural sound localization with the help of digital signal processing units. The tran-

a)Electronic mail: laurent.kneip@stud.uni-erlangen.de
b)Electronic mail: claude.baumann@education.lu

 © 2008 Acoustical Society of America

scription of the canonical rules into valuable algorithms is realized through extending powerful mathematical tools,[13] descriptions of signals and systems,[14] modern statistical and probabilistic methods,[15–17] and computational techniques like neural networks,[18,19] learning algorithms,[20] numerical analysis,[21] or predictors/estimators.[22,23] Most realized systems are intensive and costly from both the engineering and the computational point of view. Binaural models that solely rely on detectable time delay (ITD) or level difference (ILD) are supposed to be suitable for azimuthal localizations only.[24] Elevation estimates can be obtained if spectral cues are added to the considerations,[25,26] or, alternatively, if arrays of microphones are being used.[27]

By contrast, Reid and Milios[28] followed van Soest's reflections to describe the signal processing and control algorithm of a spatial sound-localizing device with two omnidirectional microphones arranged on a baseline and two rotational degrees of freedom (DOFs). The authors call their approach *active audition*, rotations of the baseline being used to identify the direction of sound as the intersection of two cones sharing the same vertex. Reid and Milios determined the common vertex by approximation since a valuable geometrical solution is not available. Although Wallach[29,30] worked out a formula that combines the elevation, the azimuth, and indirectly the interaural time difference, his equation cannot yield both unknown angles from measured ITDs, and thus does not form a complete solution to the problem (see Sec. II D). Lambert's developments[31] are not helpful here because they only concern spherical heads, while the described work does not involve any three-dimensional head, just two separated point receivers in free space.

The present analytical paper fills the gap by studying the geometrical question of the intersection of two rotationally generated consecutive cones and establishes its exact equations. It will be mathematically proved that two measurements of time differences of arrival before and after a single determined rotation of the interaural axis in the horizontal plane are sufficient to yield both the azimuth and the absolute value of the altitude (see Sec. III B). The right-left and the back-front uncertainties are clarified by this procedure, while the up-down ambiguity remains unsolved. By analogy, measuring interaural time delays before and after pivoting the interaural baseline in the frontal plane returns the altitude and the absolute value of the azimuth, keeping the back-front discrimination undefined (see Sec. III C). On the other hand, rotations around the interaural axis do not provide additional information about the azimuth or about the altitude because of the cone invariance (see Sec. III D). In the course of this paper, it will be developed that always one angle at a time can be determined independently of the distance between both microphones and the sound velocity and hence of the surrounding medium. This makes the equations interesting for aerial and underwater[32] localization applications, especially, if they rely on beamforming or matched-field-processing[33] methods.

The effects of parallax motion of the interaural axis are also analyzed in this paper. It will be shown that translations of this axis in the horizontal or frontal plane substantially contribute to the sound localization (see Sec. IV). Particularly, it will be demonstrated that with astutely combined translatory and rotational movements, all three Cartesian coordinates and thus both the three dimensional (3D) direction and the distance to the sound source are determined (see Sec. V). Furthermore, it will be proved that a shift of the binaural system on the interaural axis yields the distance and the coordinate of the sound source on that axis, even without additional rotation (see Sec. IV C).

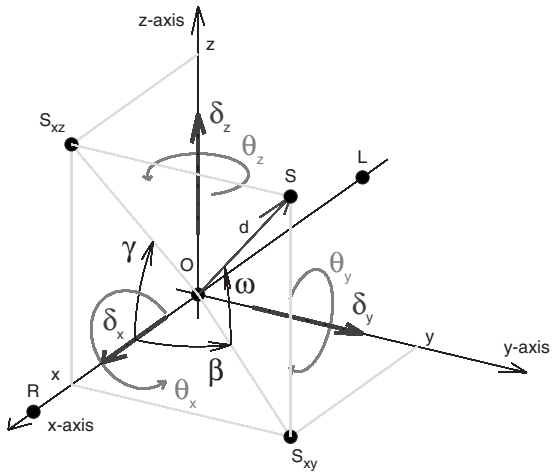


FIG. 1. The sound-source vector $\vec{OS}$ in space and the associated projections on the horizontal plane (O,x,y) yielding the azimuth β and on the frontal plane (O,x,z) returning the altitude γ. The arrows indicate the six DOFs of the system referenced to the binaural sensor (R,L) and the arbitrarily fixed orthogonal y- and z-axes. The elevation ω represents the angle between the horizontal plane and the sound-source vector (see Table I).

The presented binaural localization model follows rather elementary trigonometric reflections, contrasting with other methods that are based on more complex mathematical descriptions. It may therefore advantageously be implemented into devices with moderate computational, technical, and engineering requisites, thus reducing development and production costs (see Sec. VI).

II. FUNDAMENTALS

A. Assumptions

In the following, it is assumed that a robot[34] is equipped with an ideal binaural sensorial system capable of precisely measuring ITDs under any conditions and composed of a pair of omnidirectional microphones in the free field at a fixed distance. Techniques for estimating the ITD and the evaluation of its accuracy are not discussed here. The robot's head represents the dimensionless support of the binaural system. Only ITDs from a single immobile sound source are available. The sound origin is concentrated on a single point. There will be no scattering at the head and thus no ILDs, no pinnae, and consequently no spectral cues. The orthogonal system of coordinates (O,x,y,z) is always referenced to the head's position and orientation and thus must not be confused with the world's coordinate system. The space is subdivided as indicated in Fig. 1. Three rotational DOFs are defined for the robot head with the controlled movements: *rotating* around the z-axis, *pivoting* around the y-axis, and *tipping* around the x-axis. (The latter represents the interaural

TABLE I. Nomenclature.

Symbol	Explanation
$R(k,0,0)$	Position of the right microphone, $k>0$
$L(-k,0,0)$	Position of the left microphone
$2k$	Distance between the two microphones
$S(x,y,z)$	Sound-source location
$S_{xy}(x,y,0)$	Orthogonal projection of S on the horizontal plane
$S_{xz}(x,0,z)$	Orthogonal projection of S on the frontal plane
$\Pi(R,L,S)$	Semiplane defined by the binaural system and the sound source, characterized by the origin O, the orthogonal unitary vectors $\vec{x},\vec{v}$, and the borderline $(O,\vec{x})$
τ	Interaural time delay (ITD)
d_R,d_L,d	Distance to the sound source equal to $\|\vec{RS}\|,\|\vec{LS}\|,\|\vec{OS}\|$
$d_L-d_R=c\tau$	Path difference of arrival (c=sound velocity)
$\eta=d/k$	Relative distance
$\phi\in[0,\pi]$	Slope of the asymptote in the semiplane Π
$\varphi\in[0,\pi]$	True sound bearing in the semiplane Π
$e=\phi-\varphi$	Deviation between slope of the asymptote and true sound bearing
$\beta\in[-\pi,\pi]$	Horizontal angle (=azimuth), referenced to the x-axis
$\gamma\in[-\pi,\pi]$	Vertical angle (=altitude), referenced to the x-axis in the frontal plane
$\omega\in[-\pi,\pi]$	Elevation, vertical angle between horizontal plane and sound-source vector
$\theta_x,\theta_y,\theta_z$	Rotation angle around the specified axis
$\delta_x,\delta_y,\delta_z$.	Translation length parallel to the specified axis

axis.) By analogy three translatory DOFs are described with the controlled displacements of the robot head: *z-translation* parallel to the z-axis, etc. The nomenclature is listed in Table I. It is assumed that, before and after any defined movement, the binaural system always determines the respective ITD (τ_0 and τ_1). The sound source S is defined either by the Cartesian (x,y,z) or the polar coordinates (d,β,γ), where d represents the distance to the origin, β the azimuth, and γ the altitude, both angles being referenced to the interaural axis. It is essential to distinguish the altitude γ from the elevation ω, which denotes the angle between the horizontal plane and the sound-source vector.

B. Hyperbola of confusion in the semiplane Π

The binaural system measures the interaural time difference τ, with the sign indicating whether the sound arrives from the right or left side of the median plane (O,y,z). All potential origins of the sound event in the semiplane Π (delimited by the interaural axis and containing the sound source) with the same ITD are located on a hyperbola of confusion (Fig. 2) defined by both focuses R, L and the equation

$$\frac{x^2}{a^2}-\frac{v^2}{b^2}=1 \quad \left(a=k\cos\phi=\frac{d_L-d_R}{2}\right)\wedge(b=k\sin\phi), \tag{1}$$

with $\mathrm{sgn}(x)=\mathrm{sgn}(\tau)$, where $2k$ represents the microphone spacing and d_R, d_L the respective distances to the sound source. The asymptote of the hyperbola is described with $v=\tan\phi\cdot x$ and leads to the equation

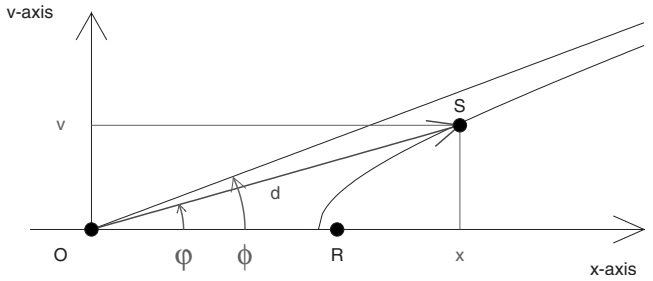


FIG. 2. The first quadrant of Π and the coordinates of the sound source S in this plane. The figure indicates the deviation in sound direction e between the asymptotically approximated direction ϕ and the true bearing φ.

$$\tau=\frac{2k}{c}\cos\phi, \tag{2}$$

which is a variant on the fundamental relation between the path difference of arrival and ϕ that has been established by Von Hornbostel and Wertheimer.[35]

C. Deviation between asymptote and hyperbola cues

The slope ϕ of the asymptote overestimates the true sound direction φ (Fig. 2). (For symmetry reasons the analysis is limited to the first quadrant.) Obviously, both angles coincide in the far field. From $d^2=x^2+v^2$ and Eq. (1) follows $x^2=\cos^2\phi\ (d^2+k^2\sin^2\phi)$. Using $\cos\varphi=x/d$, the absolute deviation e may be expressed in function of ϕ and the relative distance η,

$$e=\phi-\varphi=\phi-\arccos\left(\cos\phi\sqrt{1+\frac{1}{\eta^2}\sin^2\phi}\right). \tag{3}$$

Solving the equation $de/d\phi=0$, by transforming it to $3\cos^4\phi-(2\eta^2+3)\cos^2\phi+(\eta^2+1)=0$, proves that the deviation is maximal for the unique real root in the first quadrant

$$\phi_{e_{\max}}=\arccos\sqrt{\frac{3+2\eta^2-\sqrt{4\eta^4-3}}{6}}\approx\arccos\frac{\sqrt{2}}{2}=\frac{\pi}{4}.$$

The simplification is allowed if $4\eta^4\gg3\Leftrightarrow\eta\gg0.93$. Using Eq. (3) and $\varphi=(\pi/4)-e$, the function $\eta(e)=1/\sqrt{2\sin(2e)}$ allows to calculate that the deviation is less than 1° for sound distances greater than $4k$ and less than 0.1° for distances greater than $12k$. In the following it is therefore assumed that the sound source is not located in head proximity.

D. Cone of confusion

In order to add the third dimension to the analysis, the semiplane Π is rotated around the baseline (RL) generating a hyperboloid or asymptotically a conic hyperplane of confusion that is composed of all the points with identical time difference of arrival (Fig. 3). For any $\phi\in[0,\pi/2)\cup(\pi/2,\pi]$ the cone of confusion may be considered as an infinite set of circles with linearly growing radii and expressed with the equation

$$y^2+z^2=(x\tan\phi)^2\Leftrightarrow-\tan^2\phi\cdot x^2+y^2+z^2=0, \tag{4}$$

with $\mathrm{sgn}(x)=\mathrm{sgn}(\tau)$. If $\phi\in[0,\pi/2)$, the sound source is situated on the right side, whereas $\phi\in(\pi/2,\pi]$ describes sound

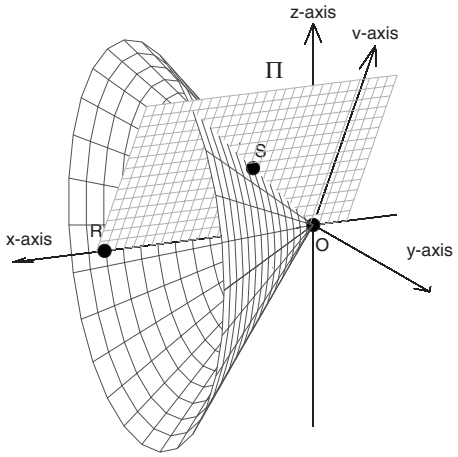


FIG. 3. The asymptotical cone generated by a rotation of Π around the interaural axis. The sound source *S* is situated on this cone.

sources on the left side of the median plane (O,y,z). The relationship between the distance and the x-coordinate is expressed with

$$x = d\cos\phi. \tag{5}$$

If $\phi=\pi/2$ then $\tau=0$ and the whole median plane defined with the equation $x=0$ represents the set of all possible sound-source locations.

Theorem A: For any measured $\tau\neq 0$, *knowing the azimuth* β *determines the absolute value of the altitude* γ *and vice versa. The relationship between* β *and* γ *is described with*

$$\tan^2\beta + \tan^2\gamma = \tan^2\phi. \tag{6}$$

Proof: From Eq. (2) it can be deduced that ϕ is unique for each τ. Furthermore, $\tau\neq 0 \Leftrightarrow \phi\neq\pi/2 \Leftrightarrow x\neq 0$. Considering any circle with radius $v=\tan\phi\cdot x$, the azimuth and the altitude may be described with $\tan\beta=y/x$ and $\tan\gamma=z/x$. The circle is defined by $y^2+z^2=v^2$, which leads to Eq. (6) and

$$(|\beta|\leq\phi)\wedge(|\gamma|\leq\phi) \quad \text{if } \phi\in\left[0,\frac{\pi}{2}\right), \tag{7a}$$

$$(|\beta|\geq\phi)\wedge(|\gamma|\geq\phi) \quad \text{if } \phi\in\left(\frac{\pi}{2},\pi\right]. \tag{7b}$$

In the special case where $\tan\phi=0$, both angles are either equal to 0 or π, depending on the sign of τ. From Eq. (6) it can be concluded that, if β is known, $|\gamma|$ may be calculated and vice versa (Q.E.D.). The elevation ω is easily calculated using

$$\tan\omega = \cos\beta\tan\gamma. \tag{8}$$

Disregarding the discussion about signs and limits, this allows rearranging Eq. (6) in order to obtain a variant on Wallach's[29,30] fundamental relationship between the azimuth β, the elevation ω, and the angle ϕ, which is related to the ITD τ through Eq. (2).

$$1+\tan^2\phi = 1+\tan^2\beta+\tan^2\gamma$$

$$\Rightarrow \frac{1}{\cos^2\phi} = \frac{1}{\cos^2\beta} + \frac{\tan^2\omega}{\cos^2\beta} = \frac{1+\tan^2\omega}{\cos^2\beta}$$

$$= \frac{1}{\cos^2\beta\cos^2\omega} \Rightarrow \cos\phi = \cos\beta\cos\omega. \tag{9}$$

III. ROTATIONS OF THE INTERAURAL AXIS

A. Transformations of the system of coordinates

Since the coordinate system always is referenced to the robot head, localization data that have been determined *before* any head movement must be considered from the new position or orientation of the reference system. For instance, the coordinates of the old cone of confusion must be transformed to the new system by isometry. This paper only analyzes "pure" movements, which are characterized by the fact that one coordinate remains invariable in the case of a rotation and two in the case of a translation. The transform equation for rotations is $\boldsymbol{X}_0=\boldsymbol{R}_\xi\cdot\boldsymbol{X}_1$ for $\xi=x,y,z$, with the matrices

$$\boldsymbol{R}_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{pmatrix},$$

$$\boldsymbol{R}_y = \begin{pmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{pmatrix},$$

$$\boldsymbol{R}_z = \begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{10}$$

B. Rotation about the *z*-axis

Theorem B. After any z-rotation of the interaural axis $\theta_z\neq 0+j\pi$, $j\in\mathbb{Z}$, *the azimuth* β_1 *is determined by*

$$\beta_1 = \arctan\left(\cot\theta_z - \frac{\tau_0}{\sin\theta_z\cdot\tau_1}\right) + i\pi, \tag{11}$$

with $i=0$ *if* $\tau_1>0$ *and* $i=1$ *if* $\tau_1<0$.

Proof: Considering the projection $\vec{OS}_{xy}$ of the sound vector $\vec{OS}$ on the horizontal plane, the azimuth in the new reference system after the z-rotation is defined with $\beta_1=\beta_0-\theta_z$. Referenced to the system of coordinates before the rotation, the line (OS_{xy}) in the horizontal plane is expressed by $y_0=\tan\beta_0\cdot x_0$, with $|\beta_0|\neq\pi/2$. Referenced to the system after the rotation, assuming that $|\beta_1|\neq\pi/2$, this changes to $y_1=\tan\beta_1\cdot x_1$, with

$$\tan\beta_1 = \frac{\sin\beta_1}{\cos\beta_1} = \frac{\sin(\beta_0-\theta_z)}{\cos\beta_1}\frac{\sin\theta_z}{\sin\theta_z}.$$

Expanding the numerator by trigonometry

$$(\sin\beta_0\cos\theta_z-\cos\beta_0\sin\theta_z)\sin\theta_z$$
$$=\sin\theta_z\sin\beta_0\cos\theta_z-(1-\cos^2\theta_z)\cos\beta_0$$
$$=\cos\theta_z(\cos\beta_0\cos\theta_z+\sin\beta_0\sin\theta_z)-\cos\beta_0$$

follows

$$\tan\beta_1=\cot\theta_z-\frac{\cos\beta_0}{\sin\theta_z\cos\beta_1}. \tag{12}$$

Since the length of the horizontal projection of the sound vector and the distance to the sound source are invariable through the rotation, the following equations are valid, assuming that $\phi_1\neq\pi/2$:

$$(x_0=\cos\phi_0\cdot d)\wedge(x_0=\cos\beta_0\cdot\|\vec{OS}_{xy}\|)$$

$$(x_1=\cos\phi_1\cdot d)\wedge(x_1=\cos\beta_1\cdot\|\vec{OS}_{xy}\|)$$
$$\Rightarrow\frac{x_0}{x_1}=\frac{\cos\phi_0}{\cos\phi_1}=\frac{\cos\beta_0}{\cos\beta_1}. \tag{13}$$

Neither $\cos\beta_0$ nor $\cos\beta_1$ are directly measurable. However, according to Eq. (2), $\cos\phi_0$ and $\cos\phi_1$ are proportional to the ITDs τ_0 and τ_1, respectively, which are yielded by the binaural sensorial system. By replacement in Eq. (12), the formula is established,

$$\tan\beta_1=\cot\theta_z-\frac{\cos\phi_0}{\sin\theta_z\cos\phi_1}. \tag{14}$$

Applying inequalities (7a) and (7b), it is concluded that

$$\beta_1=\arctan\left(\cot\theta_z-\frac{\cos\phi_0}{\sin\theta_z\cos\phi_1}\right)+i\pi \tag{15}$$

with $i=0$ if $\phi_1\in[0,\pi/2)$ and $i=1$ if $\phi_1\in(\pi/2,\pi]$. (Ignoring here the discussion about special cases where $(\phi_0=\pi/2)\wedge(\phi_1=\pi/2)$.) By replacement from Eq. (2) follows

$$\tan\beta_1=\cot\theta_z-\frac{\tau_0}{\sin\theta_z\cdot\tau_1} \tag{16}$$

and Eq. (11) (Q.E.D.). Important notice is that Eqs. (11) and (16) are independent of the microphone spacing and the sound velocity.

Finding the intersection of consecutive cones of confusion is equivalent to solving the system of the cone equations [Eq. (17b) is obtained after applying the equation $\boldsymbol{X}_0=\boldsymbol{R}_z\cdot\boldsymbol{X}_1$]

$$-\tan^2\phi_1\cdot x_1^2+y_1^2+z_1^2=0, \tag{17a}$$

$$x_1^2+y_1^2+z_1^2-\frac{1}{\cos^2\phi_0}(\cos\theta_z\cdot x_1-\sin\theta_z\cdot y_1)^2=0. \tag{17b}$$

Subtracting both equations leads to

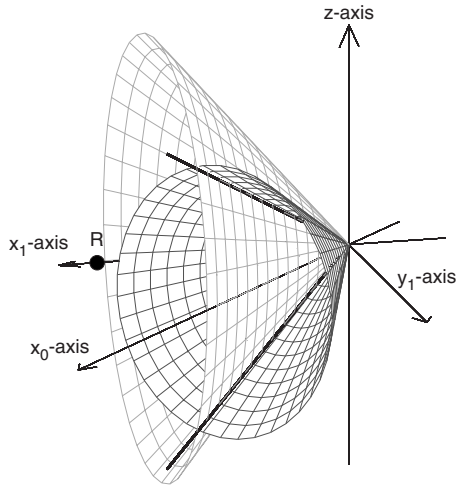


FIG. 4. The two intersectional semilines generated by a pair of cones resulting from a z-rotation. The dark cone represents a measurement before the rotation of the interaural axis.

$$\frac{1}{\cos^2\phi_0}(\cos\theta_z\cdot x_1-\sin\theta_z\cdot y_1)^2-\frac{1}{\cos^2\phi_1}x_1^2=0$$
$$\Leftrightarrow\left[y_1-\left(\cot\theta_z-\frac{\cos\phi_0}{\sin\theta_z\cos\phi_1}\right)x_1\right]$$
$$\times\left[y_1+\left(\cos\theta_z-\frac{\cos\phi_0}{\sin\theta_z\cos\phi_1}\right)x_1\right]=0.$$

From Eq. (14) and $y_1=\tan\beta_1\cdot x_1$, it is proved by contradiction that the second term of this product may not be zero for any x_1, except if $\beta_1=0+i\pi$. By replacement in Eq. (17a) and applying Theorem A follows

$$z_1=\pm\sqrt{\tan^2\phi_1-\tan^2\beta_1}\cdot x_1=\pm|\tan\gamma_1|x_1.$$

Thus, the intersection of two consecutive cones of confusion is a set of two symmetric semilines (Fig. 4). The following lemma can be formulated.

Lemma B.1. After any z-rotation of the interaural axis $\theta_z\neq0+j\pi$, $j\in\mathbb{Z}$, *the absolute value of the altitude* γ_1 *is determined.*

Conclusion: A single rotation about the z-axis determines the azimuth and the magnitude of the altitude but does not solve the up-down ambiguity.

C. Rotation about the *y*-axis

Theorem C. After any y-rotation of the interaural axis $\theta_y\neq0+j\pi$, $j\in\mathbb{Z}$, *the altitude* γ_1 *is determined by*

$$\gamma_1=\arctan\left(\cot\theta_y-\frac{\tau_0}{\sin\theta_y\cdot\tau_1}\right)+i\pi, \tag{18}$$

with $i=0$ *if* $\tau_1>0$ *and* $i=1$ *if* $\tau_1<0$.

Proof: By analogy to Theorem B, the proof of this theorem uses the equations

$$\tan\gamma_1=\cot\theta_y-\frac{\cos\phi_0}{\sin\theta_y\cos\phi_1}=\cot\theta_y-\frac{\tau_0}{\sin\theta_y\cdot\tau_1} \tag{19a}$$

$$\gamma_1 = \arctan\left(\cot\theta_y - \frac{\cos\phi_0}{\sin\theta_y \cos\phi_1}\right) + i\pi \tag{19b}$$

with $i=0$ if $\phi_1 \in [0, \pi/2)$ and $i=1$ if $\phi_1 \in (\pi/2, \pi]$. The intersection of the successive cones of confusion also is a set of two symmetric semilines.

Lemma C.1. After any y-rotation of the interaural axis $\theta_y \neq 0 + j\pi$, $j \in \mathbb{Z}$, *the absolute value of the azimuth* β_1 *is determined.*

Conclusion: A single rotation about the *y*-axis determines the altitude and the magnitude of the azimuth but does not solve the back-front ambiguity.

Note that since the (O, y, z) referential can be arbitrarily fixed in the median plane (see Fig. 1), Theorems B and C may be considered as completely equivalent. In fact, the theorems prove that one polar coordinate of the sound source is yielded in that particular plane, where the rotation of the interaural axis is executed, even if the sound source does not belong to that plane.

D. Rotation about the *x*-axis

Theorem D. Any x-rotation of the interaural axis does not provide any directional information derived from ITDs.

Proof: The equation in the matrix form of the cone of confusion before the rotation is

$$\boldsymbol{X}_0^t \cdot \boldsymbol{Q} \cdot \boldsymbol{X}_0 = 0 \quad \text{with } \boldsymbol{Q} = \begin{pmatrix} -\tan^2\phi & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{20}$$

In the new referential it is transformed to

$$(\boldsymbol{R}_x \cdot \boldsymbol{X}_1)^t \cdot \boldsymbol{Q} \cdot (\boldsymbol{R}_x \cdot \boldsymbol{X}_1) = \boldsymbol{X}_1^t \cdot \boldsymbol{R}_x^t \cdot \boldsymbol{Q} \cdot \boldsymbol{R}_x \cdot \boldsymbol{X}_1 = \boldsymbol{X}_1^t \cdot (\boldsymbol{R}_x^t \cdot \boldsymbol{Q} \cdot \boldsymbol{R}_x) \cdot \boldsymbol{X}_1 = 0.$$

By development it can easily be verified that the generated transform matrix $(\boldsymbol{R}_x^t \cdot \boldsymbol{Q} \cdot \boldsymbol{R}_x)$ is identical to $\boldsymbol{Q}$ and the cone equation remains invariable (Q.E.D.).

IV. TRANSLATIONS OF THE INTERAURAL AXIS

A. Transformations of the system of coordinates

The transform equation for translations is $\boldsymbol{X}_0 = \boldsymbol{T}_\xi + \boldsymbol{X}_1$ for $\xi = x, y, z$, with the matrices

$$\boldsymbol{T}_x = \begin{pmatrix} \delta_x \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{T}_y = \begin{pmatrix} 0 \\ \delta_y \\ 0 \end{pmatrix}, \quad \boldsymbol{T}_z = \begin{pmatrix} 0 \\ 0 \\ \delta_z \end{pmatrix}. \tag{21}$$

B. The *y*-translation and the *z*-translation

Theorem E. After any y-translation $\delta_y \neq 0$, *the projection on the horizontal plane of the intersection of the cones before and after the displacement is a subset of a parabola, which is described by*

$$y_1 = \frac{\tan^2\phi_0 - \tan^2\phi_1}{2\delta_y} x_1^2 - \frac{\delta_y}{2}, \tag{22}$$

the values of x_1 *being either positive or negative according to the sign of* τ_1.

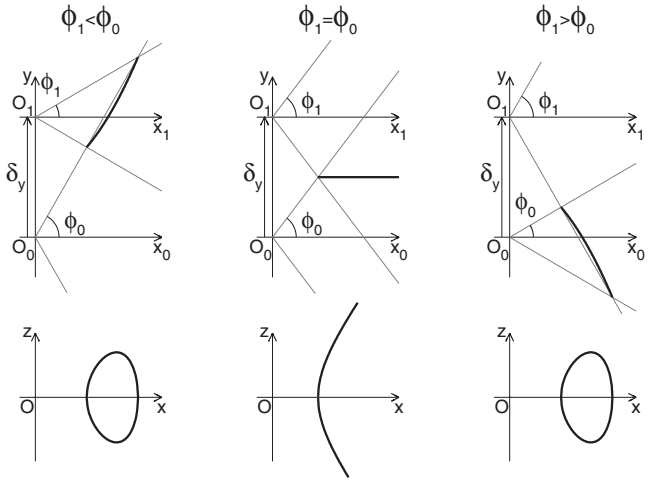


FIG. 5. The projections on the horizontal and the frontal plane of the intersection of a pair of cones obtained by *y*-translation under three possible conditions.

Proof: The *y*-coordinate in the new reference system is $y_1 = y_0 - \delta_y$. Solving the system of equations of two consecutive cones

$$-\tan^2\phi_1 \cdot x_1^2 + y_1^2 + z_1^2 = 0 \tag{23a}$$

$$-\tan^2\phi_0 \cdot x_1^2 + (y_1 + \delta_y)^2 + z_1^2 = 0 \tag{23b}$$

leads to Eq. (22), assuming that $\phi_0 \neq \pi/2$ and $\phi_1 \neq \pi/2$. By replacement in Eq. (23a), a more complex equation is composed for the projection on the frontal plane

$$z_1^2 = -\left(\frac{\tan^2\phi_0 - \tan^2\phi_1}{2\delta_y}\right)^2 x_1^4 + \left(\frac{\tan^2\phi_0 + \tan^2\phi_1}{2}\right) x_1^2 - \frac{\delta_y^2}{4}.$$

In the case of positive τ_1, in order for z_1 to be real, then x_1^2 must be between the two positive roots of the quadratic equation above (it is straightforward to show that both roots are positive). Thus, a point only belongs to the intersection of the cones if its *x*-coordinate is an element of the interval between both positive or negative roots of this fourth degree curve, or in the case $\phi_0 = \phi_1$, to the infinite interval starting from the positive or negative root, in both cases with respect to the sign of τ_1 (Q.E.D.). (Ignoring here the discussion about special cases where $\phi_0 = \pi/2$.) The different cases of intersections are shown in Fig. 5. Note that by analogy, *Theorem F* can be developed for *z*-translations.

C. The *x*-translation

Theorem G: After any x-translation $\delta_x \neq 0$, *the coordinate on the interaural axis and the distance to the sound source are determined by*

$$d_1 = \frac{x_1}{\cos\phi_1} = \frac{\sin\phi_0 \cdot \delta_x}{\sin(\phi_1 - \phi_0)}. \tag{24}$$

Proof: The *x*-coordinate in the new reference system is $x_1 = x_0 - \delta_x$. The cones of confusion that appear before and after the displacement are defined as two infinite sets of

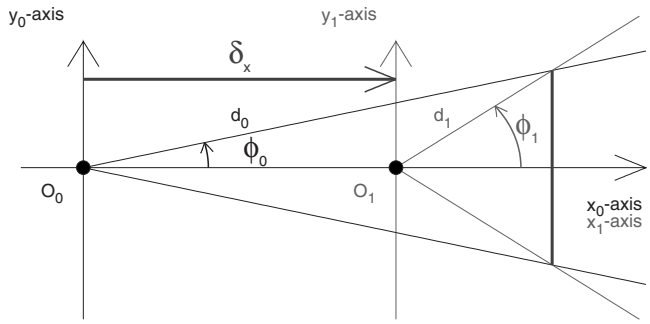


FIG. 6. The horizontal cross section of the intersection of a pair of cones obtained by x-translation.

circles with linearly growing radii. The circles all are orthogonal to the x-axis and the circle centers are located on this axis. Consequently, the intersection of the cones of confusion must be a circle, which radius fulfills the equation (proves that $\phi_0 \neq \phi_1$, assuming that $\phi_0 \neq \pi/2$ and $\phi_1 \neq \pi/2$)

$$\tan \phi_1 \cdot x_1 = \tan \phi_0 \cdot x_0 = \tan \phi_0 \cdot (x_1 + \delta_x)$$

and the x-coordinate is expressed by

$$x_1 = \frac{\tan \phi_0 \cdot \delta_x}{(\tan \phi_1 - \tan \phi_0)}. \tag{25}$$

The distance to the sound source, having changed through the movement, then is given with Eq. (24) (Q.E.D.). (Ignoring the discussion about the special cases $\phi_0 = \pi/2$ or $\phi_1 = \pi/2$.) Figure 6 shows the intersection between both cones and the horizontal plane.

V. SOUND LOCALIZATION THROUGH COMBINED MOVEMENTS OF THE INTERAURAL AXIS

This section will theoretically analyze two different robot kinematics from which the sound source can be exactly localized.

A. Finding all three coordinates, the azimuth, the altitude, and the distance through a combined motion sequence

Suppose that a robot, able to effectuate y-translations and y-rotations, obeys the following instructions (Fig. 7):

(1) measure ITD τ_0 at start, compute ϕ_0
(2) y-rotate the interaural axis by θ_y
(3) y-translate with length $\delta_y/2$
(4) measure τ_1, compute ϕ_1
(5) rotate back by $-\theta_y$ (now robot head is in the initial orientation)
(6) measure τ_2, compute $\phi_2 \rightarrow \gamma_2$ and $|\beta_2|$ are determined by application of Theorem C and Lemma C.1
(7) continue y-translation with remaining length $\delta_y/2$
(8) measure τ_3, compute ϕ_3

It will now be proved that the robot has collected all the necessary information to unambiguously localize the sound source. From Eq. (22) follows

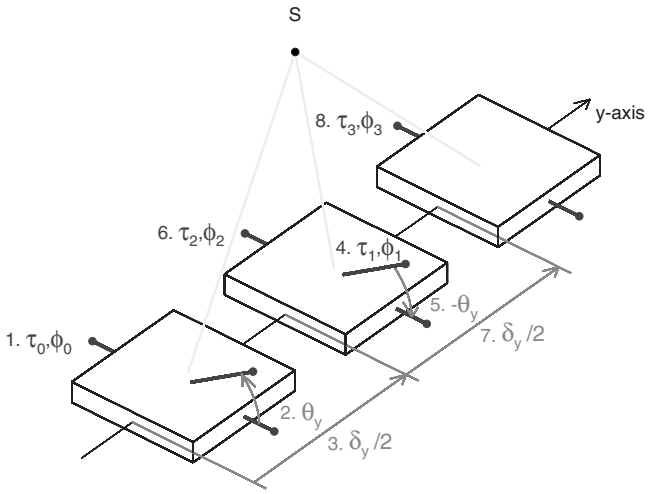


FIG. 7. Synopsis of a motion combining y-rotations and y-translations that completely solves the localization problem.

$$y_3 = \frac{\tan^2 \phi_0 - \tan^2 \phi_3}{2\delta_y} x_3^2 - \frac{\delta_y}{2}. \tag{26}$$

In the horizontal reference system (O, x_2, y_2), by definition $\tan \beta_2 = y_2/x_2$, further $x_3 = x_2 = x_0$, because of the invariance of the x-coordinate and $y_2 = y_3 + (\delta_y/2)$. Deducing by insertion,

$$x_3 = \frac{2\delta_y \tan \beta_2}{\tan^2 \phi_0 - \tan^2 \phi_3}. \tag{27}$$

The sign of β_2 is not determined at that point but the sign of x_3 is known anyway through the sign of τ_3,

$$x_3 = \mathrm{sgn}(\tau_3) \left| \frac{2\delta_y \tan \beta_2}{\tan^2 \phi_0 - \tan^2 \phi_3} \right|. \tag{28}$$

By replacement in Eq. (26) y_3 is yielded, and finally, since $z_3/x_3 = z_2/x_2 = \tan \gamma_3 = \tan \gamma_2$, $z_3 = \tan \gamma_2 \cdot x_3$ because of the invariance of the x- and z-coordinates through the translation. All coordinates being known, the distance and the direction to the sound source now are determined. Note that a similar scenario can be developed for z-translations (Theorem F).

B. Exploiting a bounded velocity differential drive robot to localize a sound source

A *bounded velocity differential drive* robot[36] has two independently driven coaxial wheels providing a single rotational DOF about the vertical axis and a translational DOF in the driving direction. If such a robot is equipped with a binaural sensor with the interaural axis pointing in the driving direction, it is able to execute z-rotations and x-translations in the binaural referential. Assuming that the system's origin O is close to the ground level, sound sources with negative altitude may be excluded and the up-down ambiguity can be considered as being solved: $\gamma = |\gamma|$. Suppose that the robot executes the following sequence of instructions (Fig. 8):

(1) measure ITD τ_0 at start
(2) z-rotate the interaural axis by θ_z (opposite spin of the wheels)

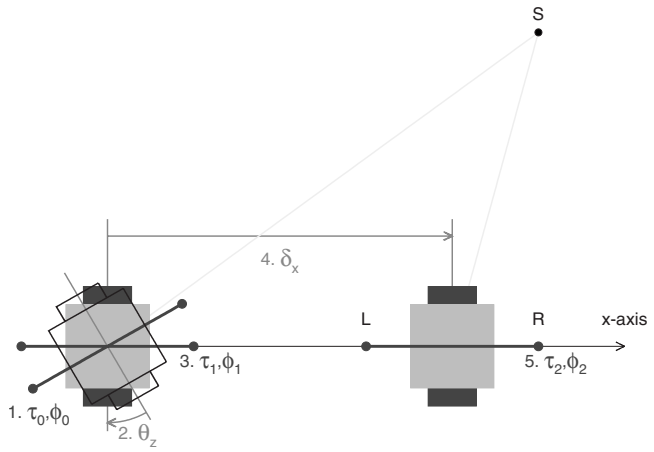


FIG. 8. Synopsis of a motion combining a single z-rotation and an x-translation that solves the localization problem.

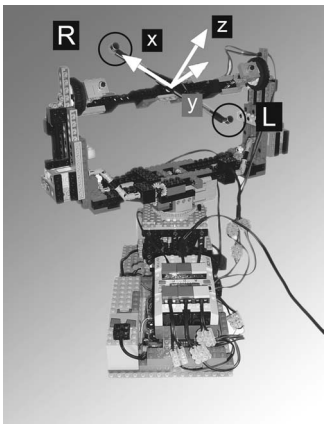


FIG. 9. The figure shows the LEGO™ robot able to fix the sound direction. The microphones (R,L) are marked on the picture. Their baseline can rotate with three DOFs.

(3) measure τ_1, compute $\phi_1 \rightarrow \beta_1$ and γ_1 are determined by application of Theorem B and Lemma B.1 (γ_1 being positive, as assumed)
(4) x-translate with length δ_x (same spin of the wheels)
(5) measure τ_2, compute $\phi_2 \rightarrow d_2$ is determined by application of Theorem G

From $x_2=\cos\phi_2 \cdot d_2$ and $x_1=x_2+\delta_x$ follows $y_2=y_1=\tan\beta_1 \cdot x_1$ and $z_2=z_1=\tan\gamma_1 \cdot x_1$. Now β_2 and γ_2 may be calculated by additional trigonometry. However, if the order of the sequence is exchanged and the z-rotation follows the x-translation, the distance to the sound source will be conserved through the second movement with the result that d_2, β_2, and γ_2 can be determined with fewer calculations. It can be concluded that after any pure rotation or translation, a differential drive robot may find all three coordinates of a sound source through only one additional complementary movement, assuming that the remaining ambiguity is solved by any means. Similar dynamic scenarios can be developed for complete and efficient sound localization in space.

Notice that the robot's mobility can be simulated through the use of specially arranged microphone arrays. For instance, one can imagine a rectangular assembly of two parallel pairs of microphones in the horizontal or frontal plane that applies the z-translation or y-translation, respectively, and simultaneously the y-rotation or z-rotation, if the pairs are considered diagonally.

VI. EXPERIMENTAL APPLICATIONS

In order to illustrate the practical value of the described theoretical binaural model, the authors realized robotic devices with the LEGO™ Mindstorms kit. Despite mechanical imprecision, this material has many advantages for prototyping because of the high variety of parts, the low costs, and the fact that robot designs are easily reproducible. The robots are equipped with two omnidirectional Electret condenser microphones and a two-channel audioamplifier forming the binaural sensor. They are controlled by the LEGO™ RCX, which essentially is a digital system based on an 8 bit microcontroller being clocked at 16 MHz. It is programed to sample 2×300 data points at 36 kHz and to determine ITDs through a fast correlation method (sum of absolute differences approach), providing an accuracy for τ of 27.78 μs. The programs use Blauert's degree of coherence[1] (p. 201) in order to prune bad cues.

A. Determining the altitude and the azimuth through *y*-rotations

The first stationary robot shown in Fig. 9 possesses all three rotational DOFs. The spacing $2k$ between both microphones pointing into the direction of the y-vector equals 16 cm. The mechanical precision of the kinematics is 3° in all three directions. The sound waves are emitted by a small music playing radio at a distance of about 1 m $[\geqslant 12k]$ from the robot's head and the tests are effectuated in an ordinary reverberant smooth-surfaced room of about 50 m^3. The sound is only presented from the front. The robot executes rotations $\theta_y=-30°$ about the y-axis. It measures the interaural time differences τ_0 before and τ_1 after the rotation. Then it calculates the altitude γ_1 and the azimuth β_1 using Eqs. (18) and (6), respectively. The exact position of the sound source is manually measured and the results are noted from the robot's display. Finally, the device turns its head in the three DOFs, in order to have the y-axis point into the direction of sound. Through this procedure the robot is able to localize the acoustic source in space with an average precision of 10°. Figure 10 shows some test results in comparison to the correct cues. The lines indicate the chronology of the test points.

In order to get an idea about the influences of the experimental variables on the localization accuracy, this paper develops a theoretical study of the errors that propagate through the system. The predicted errors are compared to the effective deviations. Errors appear in the system as ITDs are measured and rotations are produced. Rounding errors, errors in the manual measurements of the true sound-source location, and the deviations between asymptote and hyperbola cues are neglected here. The considered input and control errors are denoted with $\Delta\tau$ and $\Delta\theta_y$, respectively. The error in altitude can be estimated applying Eq. (18),

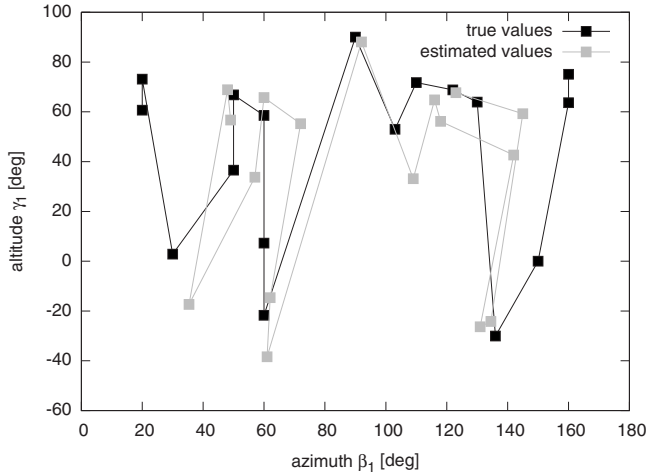


FIG. 10. A series of sound sources at altitude angles between −40° and 100° and azimuth angles between 20° and 160° have been tested.

$$\Delta\gamma_1 = \left|\frac{\partial\gamma_1}{\partial\theta_y}\right|\Delta\theta_y + \left|\frac{\partial\gamma_1}{\partial\tau_0}\right|\Delta\tau + \left|\frac{\partial\gamma_1}{\partial\tau_1}\right|\Delta\tau = \frac{|-\tau_1^2 + \tau_0\tau_1\cos\theta_y|\Delta\theta_y + (|\tau_0\sin\theta_y| + |\tau_1\sin\theta_y|)\Delta\tau}{(1+\tan^2\gamma_1)\sin^2\theta_y\cdot\tau_1^2}. \tag{29}$$

Figure 11 shows the evolution of the absolute error in altitude. The graph has been parametrized to $\tau_0/\tau_{\max}$ (expressed in percent), with $\tau_{\max}=2k/c$ and referred to the abscissa $q=\tau_1/\tau_0$, where $\tau_1\in[-\tau_{\max},\tau_{\max}]$. The shape of the curves may partly be explained from the derivate of the arctan function that is used in Eq. (18). The curve series presents significant peaks at $q\approx 1$, which must be seen in relationship with the rotation angle θ_y. In fact, assuming for a moment that the error in θ_y is 0, it can be proved that there is only one maximum, if $\theta_y\in[-2\pi/3,2\pi/3]$, whereas a second local maximum appears, if $\theta_y\in[\pi/3,5\pi/3]$. On the other hand, currently supposing that $\Delta\tau=0$, there also are two local maxima. Although the rotation angle plays an important role,

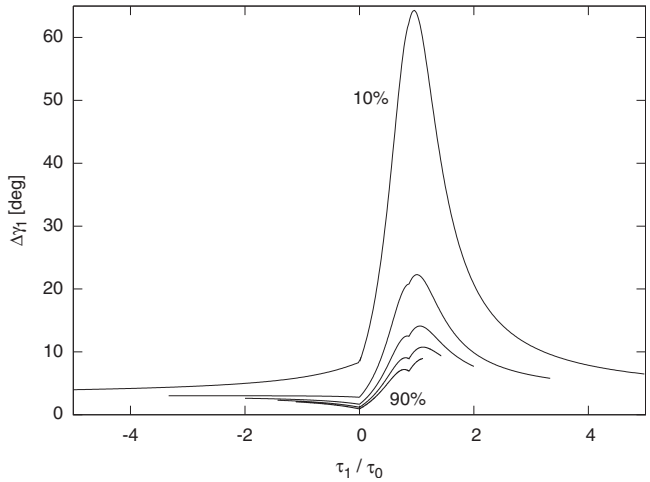


FIG. 11. Absolute error for the determination of the altitude γ_1 derived from Theorem C. The figure shows error curves for a rotation angle of $\theta_y=-30°$ as function of $q=\tau_1/\tau_0$. The parameter τ_0 varies between 0 and $\tau_{\max}$ (indicated in steps of 10%, 30%,…, 90%).

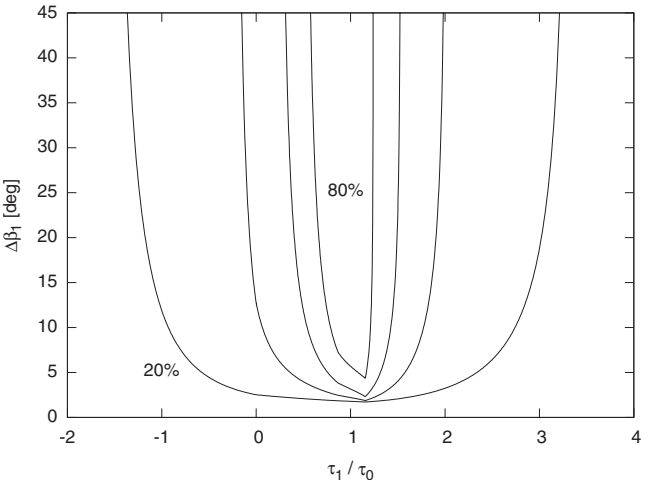


FIG. 12. Absolute error for the determination of the azimuth β_1 derived from Lemma C.1. The figure shows error curves for a rotation angle of $\theta_y=-30°$. The parameter τ_0 varies between 0 and $\tau_{\max}$ (indicated in steps of 20%, 40%,…, 80%).

the most sensitive influence on the errors in altitude comes from the measured ITDs before and after the rotation. Considering q constant, the error increases as τ_0 approaches 0 and decreases for larger τ_0, as τ_1 tends to 0^-. With

$$\Delta\phi = \left|\frac{\partial\phi}{\partial\tau}\right|\Delta\tau = \frac{\Delta\tau}{\sqrt{\left(\frac{2k}{c}\right)^2-\tau^2}}, \tag{30}$$

the expected error propagation in azimuth is described with

$$\Delta\beta_1 = \left|\frac{\partial\beta_1}{\partial\phi_1}\right|\Delta\phi_1 + \left|\frac{\partial\beta_1}{\partial\gamma_1}\right|\Delta\gamma_1 = \frac{|\tan\phi_1|(1+\tan^2\phi_1)\Delta\phi_1 + |\tan\gamma_1|(1+\tan^2\gamma_1)\Delta\gamma_1}{|\tan\beta_1|(1+\tan^2\beta_1)}. \tag{31}$$

Figure 12 shows how the error in azimuth evolves as function of q. The graph has also been parametrized to $\tau_0/\tau_{\max}$. Interestingly the curve series presents significant minima at $q\approx 1$. The region of low error becomes larger as τ_0 decreases. The error in β_1 rapidly grows as q reaches the borders of the function domain, which are expressed with $q_{a,b}=\cos\theta_y \mp \sqrt{(\tau_{\max}/\tau_0)^2-1}\cdot|\sin\theta_y|$.

It can be concluded that the application of Theorem C and Lemma C.1 may yield worst results in altitude at best azimuth estimates and vice versa. There may as well be situations, where both estimates are best. However, it does not seem possible that both angles present largest errors. It can be affirmed that the localization errors depend on the system input and control errors, the rotation angle itself, and on the spatial position of the sound source. Notice that if z-rotations had been applied, the errors in azimuth and altitude would have been inverted.

The experimental error measurements are compared to the theoretical estimates in Figs. 13 and 14. The graphs show that the real errors stay within the predicted limits. It must be underlined that the data points for each graph have been sorted according to their respective abscissa so that their or-

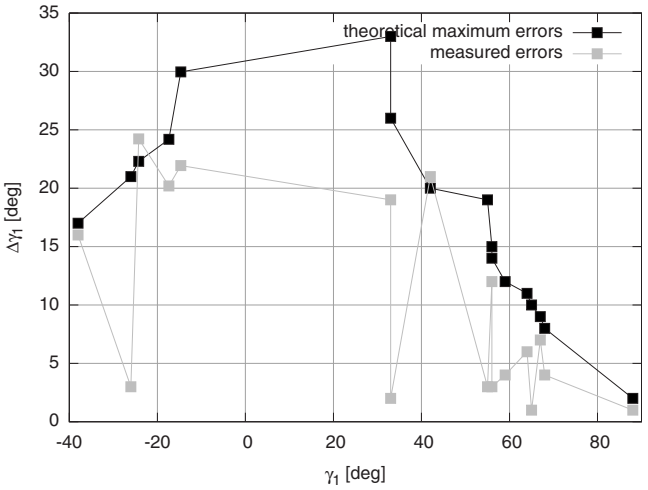


FIG. 13. Comparison between the measured and the theoretical maximum errors in the determination of the altitude γ_1 according to Eq. (29).

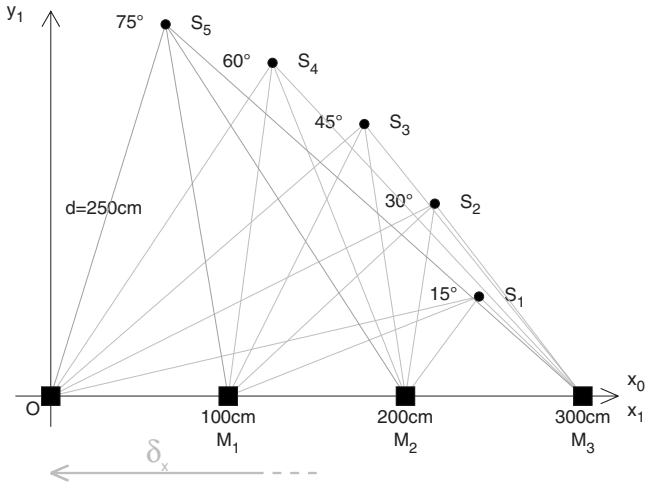


FIG. 15. The distance-measurement experiment and the tested sound-source locations, the different starting points on the x-axis, and the direction of the x-translation.

ders do no longer correspond to the one of Fig. 10. However, each data point can easily be identified on these diagrams.

The powerful mathematical model allows a simple LEGO™ robot to determine a sound direction with satisfying precision. Because each localization fix also delivers the maximum error of the angles, in future experiments the data from the model could be fed into a recursive predictor/estimator filter, improving the localization accuracy through repeated operation. Alternatively, bad cues could be pruned according to maximum error thresholds. The robot could then change the size of the rotation angle, once a bad cue being detected, and repeat the localization operation under better conditions.

B. Determining the distance through *x*-translations

The second mobile robot corresponds to the previously mentioned bounded velocity differential drive robot and serves to effectuate x-translations and to determine the distance to a sound source using Theorem G. The distance $2k$ between the two microphones now equals 21 cm. The external conditions are identical to those of the previous experiment; errors in x-translations $\Delta\delta_x$ are estimated to 2 cm and course deviations $\Delta\psi$ to 1°. Figure 15 describes the planar test scenario where several sound-source positions situated on a circle around the origin have been tried out. In each case, the robot effectuated a translation movement from three different starting points on the x-axis back to the origin, thus always trying to determine the final distance of 250 cm between the robot at the origin of the referential and the different sound-source locations. Table II shows the distance estimations for $d(O,S_i)$ with a mean square error of 46.3 cm. Using Eq. (24), the error in distance can be estimated with

$$\begin{aligned}\Delta d_1 &= \left|\frac{\partial d_1}{\partial \delta_x}\right|\Delta\delta_x + \left|\frac{\partial d_1}{\partial \phi_0}\right|\Delta\phi_0 + \left|\frac{\partial d_1}{\partial \phi_1}\right|\Delta\phi_1 + \delta_x \tan\Delta\psi \\ &= \left|\frac{\sin\phi_0}{\sin(\phi_1-\phi_0)}\right|\Delta\delta_x + \left|\frac{\delta_x \sin\phi_1}{\sin^2(\phi_1-\phi_0)}\right|\Delta\phi_0 \\ &\quad + \left|\frac{\delta_x \sin\phi_0 \cos(\phi_1-\phi_0)}{\sin^2(\phi_1-\phi_0)}\right|\Delta\phi_1 + \delta_x \tan\Delta\psi, \quad (32)\end{aligned}$$

which may be related to ITDs through Eq. (30). Figure 16 shows the corresponding curves for a translation of 2 m, parametrized to $\tau_0/\tau_{\max}$ (expressed in percent). If $q=\tau_1/\tau_0 \approx 1$, which happens if the sound source is far away from the robot, or if the displacement to distance ratio is very small, unsurprisingly, the error in distance estimation becomes very intense. Considering q constant, the error in distance increases as τ_0 decreases. In Table II it can be seen that the error in distance measurements Δd_1 mostly stays within the predicted limits.

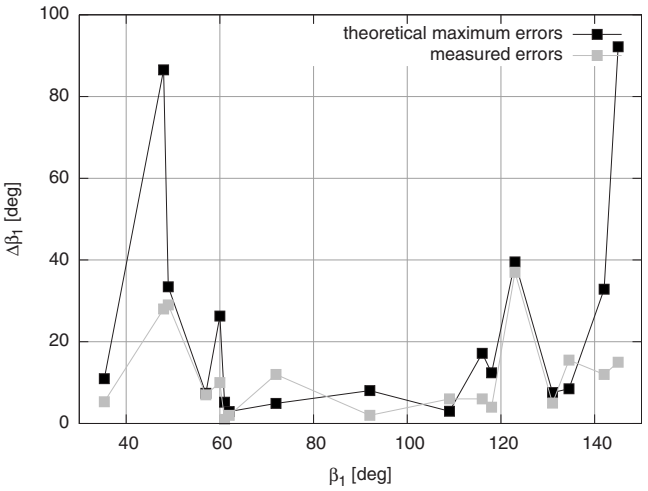


FIG. 14. Comparison between the measured and the theoretical maximum errors in the determination of the azimuth β_1 according to Eq. (31).

TABLE II. Distance estimations (in centimeters) obtained through x-translation tests (from M_i back to the origin O). The values in parentheses indicate the theoretical error in distance measurement Δd_1.

	S_1	S_2	S_3	S_4	S_5
M_1	152.5 (182)	201.6 (59)	201.2 (38)	190.7 (32)	179.1 (32)
M_2	234.8 (29)	266.9 (15)	244.5 (14)	258.6 (15)	243.0 (18)
M_3	249.1 (16)	234.1 (6)	231.6 (9)	227.9 (13)	211.0 (17)

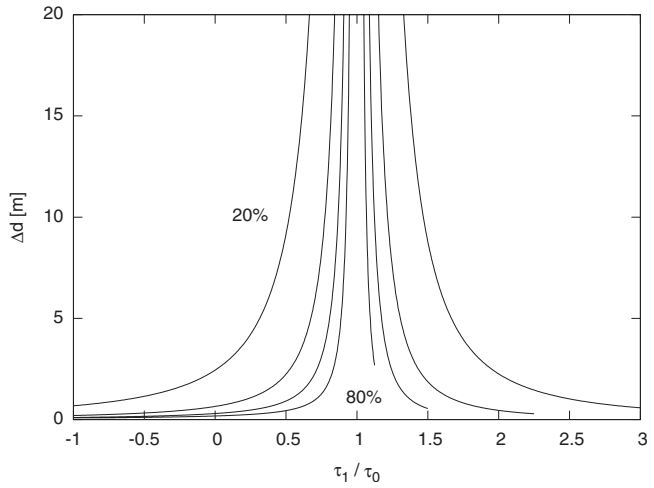


FIG. 16. Absolute error for the distance measurement according to Theorem G. The figure shows error curves for a translation of $\delta_x=2$ m. The parameter τ_0 varies between 0 and τ_{max} (indicated in steps of 20%, 40%, …, 80%).

VII. CONCLUSIONS

This paper develops a new mathematical model for efficient artificial sound localization in space. It establishes a simple formula [Eqs. (11) and (18), respectively] that determines the azimuth or the altitude of the sound source directly from the measured ITDs, as function of the effectuated rotation. This formula is independent of the microphone distance and the sound velocity and hence of the surrounding medium. This paper also proves that parallax motion unlocks additional information about the distance and the Cartesian coordinates, which, in sequential combination with a rotational movement, completely solves the localization problem. The practical value of the model is illustrated through the implementation into LEGO™ robots. The study of the error propagation through the systems shows that besides satisfying angular and distance cues, the model also delivers essential information about the accuracy of each estimation. This paper now may serve for the development of smart sound localization robots.

ACKNOWLEDGMENTS

The authors want to thank Rudolf Rabenstein, University of Erlangen-Nürnberg, Chair of Multimedia Communications and Signal Processing, for his precious help and advice. They also want to express their gratitude to Chris Rogers, Department of Mechanical Engineering, Tufts University Medford, MA, for his contribution. This article is generously supported by the Fonds National de la Recherche Luxembourg-FNR/08/MA4b/02.

[1]J. Blauert, *Spatial Hearing* (MIT Press, Cambridge, MA, 1983).
[2]J. L. van Soest, "Richtungshooren bij sinusvormige geluidstrillingen (Directional hearing of sinusoidal sound waves)," Physica (The Hague) **9**, 271–282 (1929).
[3]H. Wallach, "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol. **27**, 339–368 (1940).
[4]S. Perret and W. Noble, "The effect of head rotations on vertical sound localization," J. Acoust. Soc. Am. **102**, 2325–2332 (1997).
[5]F. Wightman and D. J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am. **105**, 2841–2853 (1999).
[6]W. R. Thurlow, J. W. Mangels, and P. S. Runge, "Head movement during sound localization," J. Acoust. Soc. Am., **42**, 489–493 (1967).
[7]J. C. Middlebrooks and D. Green, "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159 (1991).
[8]P. Coleman, "Dual role of frequency spectrum in determination of auditory distance," J. Acoust. Soc. Am. **44**, 631–632 (1968).
[9]R. A. Butler, E. T. Levy, and W. D. Neff, "Apparent distance of sounds recorded in echoic and anechoic chambers," J. Exp. Psychol. Hum. Percept. Perform. **6**, 745–750 (1980).
[10]D. H. Mershon and J. N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance," Perception **8**, 311–322 (1979).
[11]P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," J. Acoust. Soc. Am. **112**, 2110–2117 (2002).
[12]J. M. Speigle and J. M. Loomis, "Auditory distance perception by translating observers," Proceedings of the IEEE Symposium on Research Frontiers in Virtual Reality (VRAIS'93), San Jose, CA, 25–26 October 1993, pp. 92–99.
[13]W. M. Hartmann, *Signals, Sound and Sensation* (Springer-Verlag, New York, 1998).
[14]H. Schlitt, *Systemtheorie (Theory of Systems)* (Springer-Verlag, Berlin, 1992).
[15]W. M. Hartmann, B. Rakerd, and J. B. Gaalaas, "On the source identification method," J. Acoust. Soc. Am. **104**, 3546–3557 (1998).
[16]E. A. Lehmann, D. B. Ward, and R. C. Williamson, "Experimental comparison of particle filtering algorithms for acoustic source localization in a reverberant room," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), 6–10 April 2003, Vol. **5**, pp. 177–180.
[17]D. Li and Y. H. Hu, "Least square solutions of energy based acoustic source localization problems," Proceedings of the International Conference on Parallel Processing Workshops (ICPPW'04), 2004, pp. 443–446.
[18]M. S. Datum, F. Palmieri, and A. Moiseff, "An artificial neural network for sound localization using binaural cues," J. Acoust. Soc. Am. **100**, 372–383 (1996).
[19]C. Schauer, T. Zahn, P. Paschke, and H.-M. Gross, "Binaural sound localization in an artificial neural network," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'00), 2000, pp. 865–868.
[20]J. Weng and K. Y. Guentchev, "Three-dimensional sound localization from a compact noncoplanar array of microphones using tree-based learning," J. Acoust. Soc. Am. **110**, 310–323 (2001).
[21]P. Zakarauskas and M. S. Cynader, "A computational theory of spectral cue localization," J. Acoust. Soc. Am. **94**, 1323–1331 (1993).
[22]N. Strobel, S. Spors, and R. Rabenstein, "Joint audio-video object localization and tracking," IEEE Signal Process. Mag. **18**, 22–31 (2001).
[23]J. M. Valin, F. Michaud, and J. Rouat, "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering," Rob. Auton. Syst. **55**, 216–228 (2007).
[24]L. Calmes, G. Lakemeyer, and H. Wagner, "Azimuthal sound localization using coincidence of timing across frequency on a robotic platform," J. Acoust. Soc. Am. **121**, 2034–2048 (2007).
[25]C. Lim and R. O. Duda, "Estimating the azimuth and elevation of a sound source from the output of a cochlear model," Conference Record of the 28th Asilomar Conference on Signals, Systems and Computers (ACSSC'94), 1994, Vol. **1**, pp. 399–403.
[26]S. B. Andersson, A. A. Handzel, V. Shah, and P. S. Krishnaprasad, "Robot phonotaxis with dynamic sound source localization," Proceedings of the IEEE International Conference on Robotics and Automation (ICCRA'04), 26 April–1 May 2004, Vol. **5**, No. 26, pp. 4833–4838.
[27]H. Teutsch and W. Kellermann, "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," J. Acoust. Soc. Am. **120**, 2724–2736 (2006).
[28]G. L. Reid and E. Milios, "Active stereo sound localization," J. Acoust. Soc. Am. **113**, 185–193 (2003).
[29]H. Wallach, "Über die Wahrnehmung der Schallrichtung (On the perception of sound direction)," Psychol. Forsch. **22**, 238–266 (1938).
[30]H. Wallach, "On sound localization," J. Acoust. Soc. Am. **10**, 270–274 (1939).
[31]R. M. Lambert, "Dynamic theory of sound-source localization," J. Acoust. Soc. Am. **56**, 165–171 (1974).
[32]C. O. Tiemann, M. B. Porter, and J. A. Hildebrand, "Automated model-based localization of marine mammals near California (OCEANS'02)," MTS/IEEE, 2002, Vol. **3**, pp. 1360–1364.

[33]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer-Verlag, New York, 1994), Chap. 10, pp. 557–599.
[34]S. Russell and P. Norvig, *Artificial Intelligence* (Prentice-Hall, Englewood Cliffs, NJ, 2003), pp. 32–38.
[35]E. M. Von Hornbostel and M. Wertheimer, "Über die Wahrnehmung der Schallrichtung (On the perception of the direction of sound)," Sitzungsber. K. Preuss. Akad. Wiss. **20**, 388–396 (1920).
[36]D. Balkcom and M. Mason, "Time optimal trajectories for bounded velocity differential drive robots," IEEE International Conference on Robotics and Automation (ICRA'00), April, 2000, Vol. **3**, pp. 2499–2504.

Relative influence of interaural time and intensity differences on lateralization is modulated by attention to one or the other cue

Albert-Georg Lang[a] and Axel Buchner
Institut für Experimentelle Psychologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

(Received 7 May 2008; revised 4 August 2008; accepted 6 August 2008)

When setting interaural time differences and interaural intensity differences into opposition the measured trading ratio depends on which of the cues is adjusted by the listener. This paper provides some evidence that the different trading ratios may be an effect of a shift of attention toward the to-be-adjusted cue. The experiments consisted of two phases. In the compensation phase, participants canceled out the effect of one preset binaural cue by adjusting a compensatory value of the other cue until the sound was located in the center. In the localization phase participants assessed the virtual location of the sounds, again using the preset values of the fixed cue, but using the values of the other cue as previously adjusted. The sounds were no longer perceived as originating from the center. Instead, their perceived location was shifted back toward the location from which they appeared to originate before the adjustment. These findings suggest that during the compensation task the to-be-adjusted sound localization cue received an increased weight compared to the other cue. We propose shifts of attention between the cues as a mechanism that could account for this finding. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2981041]

PACS number(s): 43.66.Qp, 43.66.Ba, 43.66.Pn [JCM] Pages: 3120–3131

I. INTRODUCTION

Sound localization on the horizontal plane is largely determined by two physical parameters: interaural time differences (ITDs) and interaural intensity differences (IIDs). If experimenters or audio engineers want to manipulate the perceived location of a virtual sound source, they may specify a certain ITD, an IID, or a combination of both. The question arises which ITD has the same effect on localization as a certain IID (and vice versa). Many experiments have been conducted to find a trading ratio or an equivalence relation[1] between both interaural cues (e.g., see Leakey and Cherry, 1957, Deatherage and Hirsh, 1959, Whitworth and Jeffress, 1961, Colburn and Durlach, 1965, Hafter and Jeffress, 1968, Jeffress and McFadden, 1971, McFadden *et al.*, 1971, Hafter and Carrier, 1972, McFadden *et al.*, 1972, McFadden *et al.*, 1973, Algom *et al.*, 1988, Hafter *et al.*, 1990, Wightman and Kistler, 1992, Gaik, 1993, Wightman and Kistler, 1997, Breebaart *et al.*, 1999, Shinn Cunningham *et al.*, 2000).

Setting the binaural cues into opposition such that their effects cancel each other out is a frequently used paradigm to find an equivalence relation. An early study is that of Leakey and Cherry (1957). In this study, words spoken by a female voice were presented over two loudspeakers set up outdoors in order to avoid disturbing reflections. ITDs were implemented by placing one loudspeaker more distant from the participant than the other loudspeaker while keeping the angle between the two loudspeakers constant relative to the listener. The IIDs caused by the unequal distances between the loudspeakers and the listener were compensated by increasing the volume of the more distant speaker. Participants' task was to judge the apparent location of the virtual sound source for different combinations of ITDs and IIDs. The trials of interest were those in which the listener located the sound source exactly on the midline between the two loudspeakers. In these trials an ITD between the sound waves arriving from the two loudspeakers apparently had been compensated for by an IID of inverse sign, that is, the sound waves arrived earlier at one ear but at higher level at the other ear. It was found that the trading ratio increased with the preset ITD. For example, the ratio was 59 μs/dB with the ITD set to 450 μs, but 218 μs/dB with the ITD set to 2270 μs. However, the presentation of sounds over loudspeakers has the shortcoming that the experimenter has insufficient control over the ITDs and IIDs that arrive at the listener's ears. For this reason presentation via headphones has established as the standard procedure for ITD-IID trading experiments.

The search for an equivalence relation is complicated by experimental results that show that there are different moderator variables. For example, the ratios found by Leakey and Cherry (1957) increased with the sound level of an additionally inserted source of white noise. Deatherage and Hirsh (1959) presented low-pass-filtered clicks at 40, 60, and 80 dB_{SPL} with preset IIDs via headphones. Participants adjusted the ITD so that they located the clicks in the middle of their heads. The ratio between ITDs and IIDs was not constant over the three listening levels. The ITD required to compensate for the given IID decreased with increasing sound pressure levels. Wightman and Kistler (1992) used either wideband or high-pass-filtered noise stimuli with conflicting ITD and IID cues. In the wideband condition the listeners located the sounds solely on the basis of ITDs. Wightman and Kistler (1992) interpreted this result as evi-

[a]Electronic mail: albert.lang@uni-duesseldorf.de

 © 2008 Acoustical Society of America

dence for the dominance of low-frequency ITDs in the localization of sounds; when low-frequency ITDs were present they dominated IIDs at all frequencies.

Mills (1972) suggested that the localization judgment on the basis of the ITD limits the possible locus of a sound source to a so called "cone of confusion," that is, a cone-shaped volume of hypothetical positions of the sound source that are consistent with the perceived ITD. The IID may pose additional constraints to the location of the sound source. Shinn Cunningham *et al.* (2000) described the intersection of iso-ITD and iso-IID volumes as torus-shaped volume ("tori of confusion").

Besides the observation that the effects of opposing ITDs and IIDs cancel each other out, experiments have been reported in which setting both binaural cues into opposition leads to the perception of either of *two images*. One image is largely determined by the ITD ("time image"), while the other is determined to a much greater extent by the IID ("time/intensity image"). The trading ratios are very different: For the time image the range is about 0 to 10 μs/dB while for the time/intensity image values between 20 and 50 μs/dB have been reported. Extremely experienced listeners or participants that have completed huge amounts of practice trials over a period of several days may be able to hear both images (e.g., see Whitworth and Jeffress, 1961, Hafter and Jeffress, 1968).

Individual differences constitute another important set of moderator variables. For example, McFadden and coworkers (Jeffress and McFadden, 1971, McFadden *et al.*, 1971, McFadden *et al.*, 1972, McFadden *et al.*, 1973) used an experimental design that is known from the measurement of masking level differences. The output of a noise generator was bandpass filtered and then sent to a variable phase shifter. The fixed output (N) and the variable output (S) of the phase shifter were mixed on both earphones where S was inverted in phase on one side in relation to the other side (N0Sπ). The ITD and IID between the resultants on both ears (SN_L and SN_R) could be manipulated by adjusting the phase angle between the fixed and the variable output of the phase shifter. Among other things, they found that there were interindividual differences in participants' weighting of ITDs and IIDs when both cues were set into opposition. This was true for both detection of the signal (S) that was turned on after N and for lateralization judgments. Herman *et al.* (1977) compared the ability of older participants (aged between 60 and 72 years) with the ability of younger participants (aged between 20 and 32 years) in lateralizing a train of clicks either solely based on ITDs or solely based on IIDs. They found that the older participants needed larger ITDs to give correct lateralization judgments compared to the younger participants, but that there was no difference between the groups when IIDs were the only binaural cue.

To summarize, previous investigations of the moderators of equivalence relations concentrated on stable properties of the stimuli (overall sound level, sound level of additionally inserted noise, and dominance of low-frequency ITDs) and on traitlike interindividual differences (extreme experience, age, and individual preferences for the weighting of the cues). There are relatively few studies that investigated momentary states of the observer. Jeffress and McFadden (1971) tried to manipulate the individual weighting of ITDs and IIDs by selective feedback. They asked participants to give lateralization judgments to sound stimuli with conflicting ITD and IID information. The critical trials were preceded by practice trials in which feedback about the correctness of the lateralization judgments was either provided solely on the basis of the ITD of the two ear signals or solely on the basis of the IID. Comparing the lateralization judgments in the critical trials that followed these two types of practice trials showed no systematic differences. Thus, selective feedback did not seem to change the processing of ITDs or IIDs.

Comparisons between experiments in which participants had to compensate ITDs by the adjustment of IIDs and experiments with converse roles of both cues showed different trading ratios (for an overview, see Trahiotis and Kappauf, 1978). For example, Young and Levine (1977) conducted two experiments in which the position of a pure tone of 250, 500, or 1000 Hz had to be aligned to the position of a diotic noise marker. The test tones were either presented with a fixed ITD or with a fixed IID. Participants' task was to adjust the other binaural cue so that the perceived position of the test tone was identical to that of the noise marker. The trading ratios more strongly favored the IID when IIDs had to be adjusted in order to compensate for ITDs than in the converse case. For example, with the test tone of 500 Hz the trading ratios were 79.4 and 40.4 μs/dB, respectively. Young and Levine (1977) interpreted this result as evidence for two-image lateralization. Trahiotis and Kappauf (1978) disagreed with the assumption that the different trading ratios obtained by Young and Levine (1977) could be accounted for by a time image and a time/intensity image. As mentioned above, the time image is almost solely determined by the ITD and only to a very small degree by the IID, which leads to very small trading ratios (e.g., 3 μs/dB as reported by Whitworth and Jeffress, 1961). Since the trading ratios found by Young and Levine (1977) were relatively large in *both* experiments, Trahiotis and Kappauf (1978) considered both ratios as more characteristic of time/intensity images. They proposed that the data could be described in terms of a "regression" to the features of the reference stimulus. In order to clarify the regression interpretation Trahiotis and Kappauf (1978) discussed a study by Sheldon (1973) in which participants were asked to adjust the abruptness or "surge" of a vibrotactile stimulus in relation to a reference stimulus. The abruptness was determined by two physical parameters: the rise time and the final amplitude. The test stimulus differed from the reference stimulus either in rise time or in amplitude. When the stimuli differed in rise time, participants had to adjust the amplitude until the stimuli had the same perceived abruptness; when the stimuli differed in amplitude the rise time was the parameter to be adjusted. It was found that the "equal-surge contours" were different in both conditions; when the amplitude was to be adjusted the chosen amplitude was closer to the amplitude of the reference stimulus than predicted by the contour of the condition where the rise time was the to-be-adjusted parameter. Analogous results were found when the rise time was to be adjusted. Trahiotis and

Kappauf (1978) assumed that an analogous effect occurred in the study by Young and Levine (1977) when subjects were asked to adjust one binaural localization cue such that a test tone was localized at the same position as a reference tone. In the case of the study reported by Young and Levine (1977) the reference stimulus was presented diotically (i.e., with an ITD of 0 μs and an IID of 0 dB). Consider a trial in which the test stimulus was presented with a fixed ITD and the participant had to adjust the IID to match the location of the reference tone; since the IID of the reference tone was 0 dB, a regression to the features of the reference stimulus would mean that the chosen compensatory IID was shifted toward 0 dB, that is, its absolute value was reduced. If the roles of ITD and IID were interchanged an analog regression toward an ITD of 0 μs would occur. Trahiotis and Kappauf (1978) were clear about the fact that the term regression only describes the data they discussed, but that it does not explain the observed phenomena in terms of the cognitive processes involved.

We suggest shifts of attention as a possible explanation of equivalence relations that differ as a function of which interaural cue is adjusted. Consider that while adjusting one binaural cue in order to compensate for the other cue, participants move a control element and simultaneously receive feedback about the effects of their manipulations in terms of an immediate change in the virtual location of the sound source. Given that participants are instructed to find an adjustment value that leads to a certain localization (e.g., in a central position), they must carefully observe the relation between their adjustments and the perceived change in the location of the sound source. This could lead to an increased attention to the effect of the to-be-adjusted cue on perceived location which in turn could lead to an increased perceptual weight of this cue in relation to the other cue.

If this assumption were correct, then one would expect differences in perceived sound source location between compensation and localization tasks even for identical pairs of ITDs and IIDs. In the compensation task, attention is focused on the to-be-adjusted cue (e.g., the IID), amplifying the weight of this cue relative to the fixed cue (e.g., the ITD). In the localization task, attention is distributed more evenly between both binaural cues. Then the previously chosen compensatory value (e.g., of the IID) would no longer receive the amplified weight and thus the image would no longer appear centered; instead it would be shifted away from the adjusted cue (IID) and toward the preset cue (ITD). Furthermore, if attention shifts induced by the compensation task caused the observed differences in trading ratios depending on which binaural cue is adjusted, these trading ratio differences are also expected to occur if no reference stimulus is presented that can act as an "anchor" for regression.

Four experiments were conducted to test these predictions. In experiments 1a–1c IIDs were adjusted in order to compensate for ITDs. In experiment 2 ITDs were adjusted in order to compensate for IIDs. Each experiment consisted of two phases: During the first phase ("compensation phase") the adjustment of the IID or ITD took place; the second phase ("localization phase") consisted of pure localization judgments.

II. EXPERIMENT 1a

A. Method

1. Participants

Participants were 12 female and 4 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. Their age ranged from 24 to 45 years ($M=28$, SD $=8.3$). All participants reported normal hearing. They were paid for participating or received course credit.

2. Apparatus and stimuli

Five meaningless words ("adaxal," "eferel," "inibis," "odobol," and "urubut," pronounced according to German pronunciation rules) were spoken by a female speaker and recorded in an anechoic chamber in order to avoid disturbing reflections which would have posed additional cues for spatial hearing. The signal of the microphone (Sennheiser MD-421) was amplified by one channel of a low-noise mixing console and recorded by a digital audio tape (DAT) recorder.

In order to familiarize participants with sound localization via headphones, a special set of stimuli was created. Head-related impulse responses (HRIRs) were recorded in order to be able to present practice trials in which the five sounds had natural ITD to IID ratios. The HRIRs were recorded for lateral angles between −90° and 90° in steps of 10°.

The stimuli were presented by a program written in C++ on an Intel-based personal computer. The Microsoft "Direct Sound" programming interface was used for playing the sounds and for manipulating ITDs and IIDs of the nonpractice trials. In order to assess the accuracy of Direct Sound when delaying one channel in relation to the other, the delay times were measured in the range relevant for the experiment (−600–600 μs) using a two-channel oscilloscope. This measurement revealed that Direct Sound changes the delay in steps of integer samples (22.68 μs at 44.1 kHz).

When Direct Sound receives the instruction to change the relative level between the left and the right channel, it attenuates the weaker channel so that the desired ratio is obtained. By this approach the overall sound level varies against the chosen intensity difference. In order to keep the overall sound level constant, the level was corrected by a modified version of a formula proposed by Gaik (1993) such that

$$L_{\mathrm{L}} = 20 \log \frac{10^{(L_{\mathrm{ref}}+6\mathrm{dB})/20}}{1+10^{\Delta L/20}} \tag{1}$$

and

$$L_{\mathrm{R}} = L_{\mathrm{L}} + \Delta L, \tag{2}$$

where L_{ref} denotes the reference level, that is, the level of a sound with no intensity difference between the channels that has the same perceived loudness as the actual sound, and L_{L} and L_{R} denote the left and right sound levels, respectively; ΔL is the difference between the left and right sound levels.

During the experiment, the sounds were presented via headphones (AKG K-400) at a sound level of about 60 $\mathrm{dB_{SPL}}$ (A weighted). The ITD was varied in seven steps: −600,

−400, −200, 0, 200, 400, and 600 μs.[2] There were two types of (nonpractice) trials: During the critical trials participants had to assess the direction of a sound whose ITD they had compensated for by a certain IID in an earlier phase of the experiment. The control trials were identical to the critical trials, but their ITD had not been compensated for by the participants, that is, the IID was 0 dB. Each of the five words was presented once with each of the ITDs, resulting in 35 critical trials and 35 control trials.

3. Procedure

The experiment took place in an anechoic chamber. In an initial *demonstration phase*, the computer monitor in front of the participants showed a sketch of a head wearing headphones as seen from behind. The recorded voice of a narrator spoke to the participants with the recordings of the voice being convolved with the same impulse responses as the words of the subsequent practice trials. A red dot in the drawing showed the "actual location" of the voice, that is, it appeared at an angle relative to the head sketch that matched the lateral angle employed during HRIR recordings. Different locations were demonstrated while the narrator spoke sentences such as "It sounds like this when I talk to you from the rightmost location." Subsequently, the narrator asked the participants to use the computer mouse to move the red dot to the correct position while he said "From where do you hear me now?" This sentence was repeated with 1000 ms of silence between two repetitions. After participants had made a decision and clicked on a button labeled "continue," the playback of the sentence was stopped and a green dot indicating the correct location was displayed for 2000 ms.

The demonstration was followed by a *practice phase* of 20 trials, each of which started with playing one of the five meaningless words in a loop with 500 ms of silence between two repetitions. The lateral angle of the virtual sound source was chosen randomly (between −90° and 90° in steps of 10°). As in the demonstration, participants' task was to move the red dot to the correct location and subsequently to press the continue button after which the green dot indicating the correct position was displayed for 2000 ms. 500 ms (±100 ms random variation) after the green dot had disappeared the next trial started. Every ten trials participants were offered to take a break for as long as they wished. After the second block of practice trials performance was evaluated for the last ten trials. A location judgment was classified as incorrect if the angular deviation from the correct location was larger than 22.5°. If four or more trials were incorrect, then ten more trials were presented together with the request to concentrate on giving more precise judgments. This was repeated if necessary.

For the next two phases of the experiment (compensation phase and localization phase) unfiltered sounds were used, that is, no HRIR filter was used here; the ITDs and IIDs were manipulated directly with Direct Sound as described above.

The *compensation phase* consisted of 35 trials. Each trial started with the continuous playback of one of the five meaningless words in a loop with 500 ms of silence between two word repetitions. The sound on one ear was delayed in relation to the sound on the other ear by −600, −400, −200, 0, 200, 400, or 600 μs. The computer monitor displayed a vertical slider that was used to control the relative sound level of the left and the right channels within a range of ±15 dB. The mapping of the slider's position (top versus bottom) onto the sign of the level ratio (higher level on the left versus on the right channel) was counterbalanced across participants. The slider's initial position was chosen randomly for each trial. Participants were instructed to move the slider so that they localized the sound source in the center. Participants received no further information about the physical properties of the sound or the physical property that was changed by the slider. 500 ms (±100 ms random variation) after the continue button had been clicked the next trial started. Participants were asked to take a rest in intervals of ten trials.

During the *localization phase* each trial again started with the continuous playback of one of the five meaningless words in a loop with 500 ms of silence between two word repetitions. Participants assessed the apparent sound location of (a) the previously presented trials for which participants had compensated the preset ITD by an IID such that the sound appeared to originate from a central location (critical trials; the IIDs were constant at the final value selected by the participant during the compensation phase such that the sound appeared to originate from a central location) and (b) trials with only ITDs but no IIDs (control trials). On the monitor the same drawing of a head and red dot was presented as in the demonstration and practice phasess. The 70 localization trials (35 critical and 35 control trials) were set up in a new randomly chosen sequence. The participants' task was to move the red dot to the corresponding position at which they located the sound. All other aspects of the localization trials (such as intertrial and break intervals) were parallel to those of the compensation trials.

4. Design

The independent variable was the ITD, which was manipulated within subject in seven steps (−600, −400, −200, 0, 200, 400, and 600 μs). The dependent variables were (a) the chosen IID during the compensation phase and (b) the perceived location during the localization phase of the experiment. A multivariate approach (MANOVA) was used for the within-participant comparisons. Polynomial contrasts were evaluated from orders 1 to 4. Partial η^2 is reported as an effect size measure. A statistical power analysis using *G*Power3 (Faul *et al.*, 2007) showed that in order to detect effects of $\eta^2=0.75$ (determined in a pilot study) of the independent variable given $\alpha=\beta=0.05$, $N=9$ participants were needed.

B. Results

Figure 1 (left panel) illustrates the relation between the preset ITD and the IID chosen to compensate for the effects of the ITD during the compensation phase. A MANOVA showed that the obvious effect of the preset ITD (−600, −400, −200, 0, 200, 400, and 600 μs) on the chosen IID was statistically significant $[F(6,10)=15.25, p<0.001,$

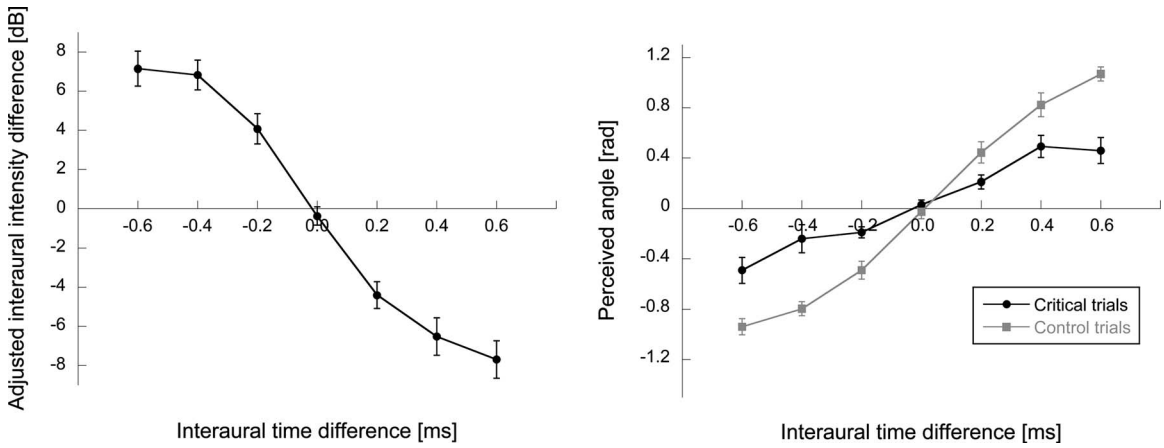


FIG. 1. Left panel: IIDs chosen to compensate for given ITDs during the compensation phase of experiment 1a (error bars denote the standard errors of the means). Right panel: Relation between the preset ITD and the perceived location during the critical trials and the control trials of the localization phase of experiment 1a (error bars denote the standard errors of the means).

$\eta^2=0.90$]. An analysis of the polynomial contrasts revealed statistically significant first and third order trends [$F(1,15)=87.13$, $p<0.001$, $\eta^2=0.85$; $F(1,15)=26.58$, $p<0.001$, $\eta^2=0.64$, respectively].

Figure 1 (right panel) shows the relation between the ITD and the perceived location during the localization phase. A MANOVA for the control trials showed a significant effect of the ITD on the perceived sound source location [$F(6,10)=49.05$, $p<0.001$, $\eta^2=0.97$].

A more interesting analysis concerns the perceived sound source locations for the critical trials. For these trials participants had compensated ITDs by appropriate IIDs during the compensation phase. Thus, one should expect, for all levels of the ITD variable, a perceived location at the midline (i.e., a perceived angle of zero) if the processes operating during the compensation and localization phase were identical. However, this does not appear to be the case. Instead, the perceived location of the critical trials is shifted toward the location predicted by the ITD, albeit not completely. A MANOVA for the critical trials showed that the effect of the ITD was indeed significant [$F(6,10)=9.63$, $p=0.001$, $\eta^2=0.85$]. An analysis of the polynomial contrasts showed that only the linear component was statistically significant [$F(1,15)=37.31$, $p<0.001$, $\eta^2=0.71$].

C. Discussion

The analysis of the compensation phase data showed that participants had understood the task of canceling out the preset ITDs by appropriate IIDs.

However, the data from the localization phase showed that the processes operating while manipulating IIDs to compensate for ITDs cannot be considered equivalent to the processes operating when sounds are located in a normal listening situation. The same sounds that had been adjusted to be perceived as originating from the midline were perceived off midline and, on average, shifted toward the location predicted by the ITDs during the localization phase. This shift-back effect is consistent with the assumption that when participants compensate for ITDs by manipulating IIDs, attention is focused on the (to-be-manipulated) IIDs. The result is that more weight is given to the IIDs than would be the case in a more neutral listening situation. During the localization phase, attention is distributed more evenly across the binaural cues and thus, the adjusted IID is not large enough to compensate for the ITD.

However, an alternative explanation for these findings is possible based on the fact that during the localization phase ITDs were over-represented. The critical trials were presented with a combination of ITDs and IIDs whereas the control trials were implemented only in terms of ITDs (i.e., the IID was 0 dB). This over-representation of ITDs could have led to an increased weighting of this cue. In order to test this hypothesis, experiment 1b was conducted with additional control trials that had only IIDs while ITDs were 0 μs.

III. EXPERIMENT 1b

A. Method

1. Participants

Participants were 12 female and 9 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. Their age ranged from 18 to 40 years ($M=27$, SD $=7.4$). All participants reported normal hearing. They were paid for participating or received course credit.

2. Apparatus and stimuli

Apparatus and stimuli were identical to those of experiment 1a except for the following: In addition to the 35 control trials that only had ITDs (−600, −400, −200, 0, 200, 400, and 600 μs), further 35 additional trials were inserted that only had IIDs (−7.5, −5.0, −2.5, 0, 2.5, 5.0, and 7.5 dB).

3. Procedure

The procedure was identical to that of experiment 1a.

4. Design

The design was identical to that of experiment 1a except for the additional independent variable of IIDs for the added control trials.

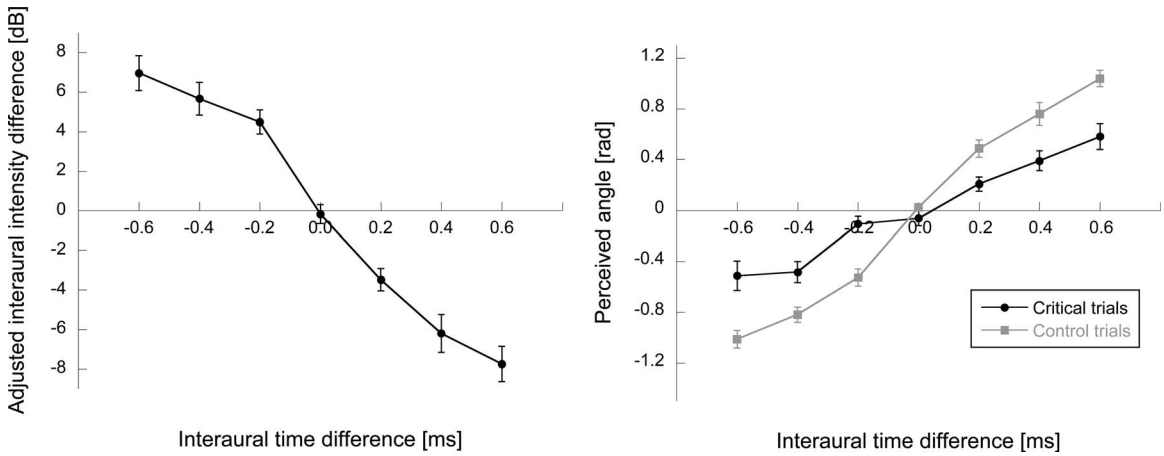


FIG. 2. Left panel: IIDs chosen to compensate for given ITDs during the compensation phase of experiment 1b (error bars denote the standard errors of the means). Right panel: Relation between the preset ITD and the perceived location during the critical trials and the (ITD) control trials of the localization phase of experiment 1b (error bars denote the standard errors of the means).

B. Results

Figure 2 (left panel) illustrates the relation between the preset ITD and the IID chosen to compensate for the effects of the ITD during the compensation phase. A MANOVA showed that the effect of the preset ITD on the chosen IID was statistically significant [$F(6,15)=18.63$, $p<0.001$, $\eta^2=0.88$].

An analysis of the polynomial contrasts showed statistically significant trends of orders 1–3 [$F(1,20)=73.76$, $p<0.001$, $\eta^2=0.79$; $F(1,20)=8.88$, $p=0.007$, $\eta^2=0.31$; $F(1,20)=15.89$, $p=0.001$, $\eta^2=0.44$].

Figure 2 (right panel) shows the relation between the ITD and the perceived location during the localization phase. A MANOVA for the (ITD) control trials showed a significant effect of the ITD on the perceived sound source location [$F(6,15)=58.86$, $p<0.001$, $\eta^2=0.96$]. A MANOVA for the control trials that had only IIDs (Fig. 3) revealed a significant effect of the IID on the perceived location of the sound source [$F(6,15)=14.94$, $p<0.001$, $\eta^2=0.86$].

A MANOVA for the critical trials showed a statistically significant effect of the ITD [$F(6,15)=7.93$, $p=0.001$, $\eta^2=0.76$]. An analysis of the polynomial contrasts revealed that only the linear component was statistically significant [$F(1,20)=44.14$, $p<0.001$, $\eta^2=0.69$].

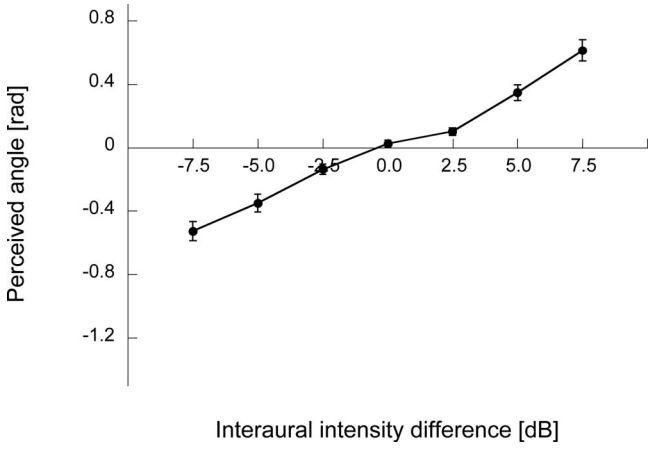


FIG. 3. Relation between the preset IID and the perceived location during the (IID) control trials of the localization phase of experiment 1b (error bars denote the standard errors of the means).

C. Discussion

The effect of the ITD on the perceived sound source location in the critical trials is similar to that of experiment 1a ($\eta^2=0.85$ in experiment 1a, $\eta^2=0.76$ in experiment 1b). The same is true for the effect size of the linear component revealed by the polynomial contrast analyses ($\eta^2=0.71$ and $\eta^2=0.69$, respectively). Hence, the fact that there were only control trials with ITDs in the localization phase of experiment 1a seems to be irrelevant for the localization judgments during the critical trials.

Differentially adapting processors for ITDs and IIDs might pose a further alternative explanation of the shift-back effect found in experiments 1a and 1b. Phillips and Hall (2005) asked participants to give laterality judgments for tone pulses of two different frequencies which differed either in ITD or in IID. The tone pulses were preceded by adaptor tones of the same two frequencies that were highly lateralized on opposite sides based on the same interaural cue as the test tones. An adaptation effect was found, that is, for each of the two test tones the perceptual weight given to the interaural cue that was present with the adaptor tone was reduced and hence the point of perceived centrality was displaced toward the side of the adaptor tone with the same frequency. In the compensation phases of the current experiments the value of the ITD was constant during each trial while the IID was regulated by the participant and hence subject to permanent change. This presentation of a fixed ITD could have led to a similar adaptation effect as described by Phillips and Hall (2005) and thus to a decrease in the perceptual weight of the ITD.

Experiment 1c was conducted in order to test this adaptation hypothesis against the attentional modulation hypothesis. In the localization phase of experiment 1a, each trial started with the continuous playback of one of the five meaningless words with constant values for the ITDs and IIDs; in

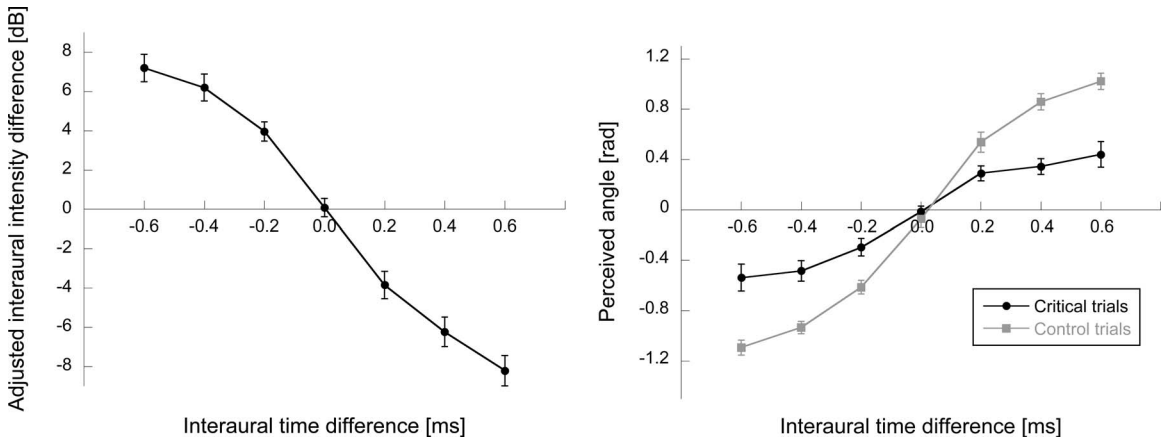


FIG. 4. Left panel: IIDs chosen to compensate for given ITDs during the compensation phase of experiment 1c (error bars denote the standard errors of the means). Right panel: Relation between the preset ITD and the perceived location during the critical trials and the control trials of the localization phase of experiment 1c (error bars denote the standard errors of the means).

the critical trials the ITDs were preset and the IIDs corresponded to the final value selected by the participant such that the sound appeared to originate from a central location in the compensation phase. In the localization phase of experiment 1c, in contrast, the critical trials in the localization phase were simply playbacks of the compensation trials. Thus, participants listened to exactly the same sequence of ITDs and IIDs as during the compensation phase, which means that during one trial, the ITD was constant at the preset value and the IID corresponded to the various values selected by the participant during the compensation phase.

Thus, the possible effect of adapting to a fixed ITD during one trial on the one hand and the effect of allocating attention to the to-be-adjusted IID on the other hand were distinguishable. If the shift-back effect found in experiments 1a and 1b were due to an adaptation to ITDs, then this adaptation must be present during both the compensation phase and the localization phase of experiment 1c because participants listen to exactly the same stimuli in the relevant trials in both phases. Hence, the sounds that were localized at a central position in the compensation phase should also be localized at a central position in the localization phase. Thus, if the adaptation hypothesis were correct, then the results of experiment 1c should deviate from those of experiments 1a and 1b in that there should be no shift-back effect in the localization phase. In contrast, if the shift-back effect found in experiments 1a and 1b were caused by shifting listeners' attention toward the IIDs by the process of regulating this cue during the compensation phase, then an off-center localization during the localization phase in which attention is more evenly distributed should still be observed. In other words, if the attentional modulation hypothesis were correct, then there should be a shift-back effect in experiment 1c that is similar to the shift-back effects observed in experiments 1a and 1b.

IV. EXPERIMENT 1c

A. Method

1. Participants

Participants were 12 female and 8 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. Their age ranged from 19 to 50 years ($M=26$, SD $=8.5$). All participants reported normal hearing. They were paid for participating or received course credit.

2. Apparatus and stimuli

Apparatus and stimuli were identical to those of experiment 1a except for the following. For every trial during the compensation phase, the entire sequence of slider movements (and, hence, the entire sequence of the participant's IID manipulations) was recorded in terms of both the movement velocity and duration (with a resolution of 1 ms) and the new position of the slider. For the critical trials of the localization phase the movements of the slider were simulated by changing the IID as recorded during the compensation phase. The IIDs of the control trials were changed according to a sine-shaped function with random amplitude (between 5 and 10 dB) and random cycle duration (between 2000 and 5000 ms). Before and after the sine-shaped change in the IID of the control trials the IID was constant at 0 dB for a random interval (between 1000 and 2000 ms before and between 2000 and 3000 ms after the changing process).

3. Procedure

The procedure was identical to that of experiment 1a except for the localization phase. Participants were told that they would hear sounds that would change in perceived direction. They were instructed to wait until the playing of the sound stopped and to indicate the final location of the sound. They saw the same sketch of the head and red dot as during the localization phase of experiment 1a and did not see the slider.

4. Design

The design was identical to that of experiment 1a.

B. Results

Figure 4 (left panel) illustrates the relation between the preset ITD and the IID chosen to compensate for the effects of the ITD during the compensation phase. A MANOVA showed that the effect of the preset ITD on the chosen IID

was statistically significant [$F(6,14)=21.41$, $p<0.001$, $\eta^2=0.90$]. An analysis of the polynomial contrasts showed statistically significant trends of orders 1 and 3 [$F(1,19)=127.50$, $p<0.001$, $\eta^2=0.87$; $F(1,19)=18.61$, $p<0.001$, $\eta^2=0.50$].

Figure 4 (right panel) shows the relation between the ITD and the perceived location during the localization phase. A MANOVA for the control trials showed a significant effect of the ITD on the perceived sound source location [$F(6,14)=94.55$, $p<0.001$, $\eta^2=0.98$]. A MANOVA for the critical trials showed a statistically significant effect of the ITD [$F(6,14)=8.36$, $p=0.001$, $\eta^2=0.78$]. An analysis of the polynomial contrasts revealed that the linear and the cubic component were statistically significant [$F(1,19)=51.84$, $p<0.001$, $\eta^2=0.73$; $F(1,19)=5.05$, $p=0.037$, $\eta^2=0.21$].

C. Discussion

The effect of ITD on the perceived sound source location in the critical trials is similar to that of experiment 1a ($\eta^2=0.85$ in experiment 1a, $\eta^2=0.78$ in experiment 1c). The linear component revealed by the polynomial contrast analyses is also similar ($\eta^2=0.71$ in experiment 1a and $\eta^2=0.73$ in experiment 1c).

The fact that the localization of the critical trials during the localization phase was off center, even though participants during the localization phase listened to the same fixed ITD and identical changes of the IID as during the compensation phase, is incompatible with the hypothesis that the shift-back effect obtained in experiments 1a and 1b was due to selective adaptation to ITDs during the compensation phase. If adaptation caused the shift-back effect in those experiments, then the effect should have disappeared in the localization phase of experiment 1c. This is so because participants listened to the same sequences of constant ITDs and changing IIDs during the compensation and the localization phase, as a consequence of which adaptation should have occurred in both phases of the experiment. In contrast, the present results are clearly compatible with the assumption that the *task of adjusting* the IID shifted listeners' attention toward the IIDs, thus giving more perceptual weight to this cue during the compensation phase than during the localization phase.

An analogous shift-back effect as in experiments 1a–1c should be observed when the roles of ITDs and IIDs are reversed. Experiment 2 was conducted to test this prediction. However, as mentioned in the Introduction, there is some evidence that ITDs play a dominant role in localization when low-frequency components are contained in the sounds (Wightman and Kistler, 1992, Shinn Cunningham *et al.*, 2000). If this were correct, then during the compensation phase we would try to shift participants' attention toward the interaural cue that already is the dominant cue. This could lead to a decrease in the shift-back effect. To anticipate, this is what we found.

V. EXPERIMENT 2

A. Method

1. Participants

Participants were 29 female and 10 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. Their age ranged from 18 to 47 years ($M=24$, SD=5.8). All participants reported normal hearing. They were paid for participating or received course credit.

2. Apparatus and stimuli

Apparatus and stimuli were identical to those of experiment 1a, with the following exceptions. The sounds were presented with preset IIDs instead of ITDs. As a reference point for choosing a reasonable range of IIDs, the mean value of the IIDs chosen by the participants to compensate for the maximum ITD (± 600 μs) of experiment 1a was computed. The mean IID for this judgment situation was 7.42 dB. We therefore selected ITDs of −7.5, −5.0, −2.5, 0, 2.5, 5.0, and 7.5 dB for experiment 2. The slider in the compensation phase of the experiment allowed participants to adjust the ITD in a range of ± 600 μs.

During a pilot study, Lang had the subjective impression that in some trials even an ITD of ± 600 μs would not seem to compensate for the given IID. For this reason a checkbox labeled "not enough" was displayed next to the slider. Participants were instructed to check the box if they had the impression that even the most extreme slider position was not sufficient to achieve a sound localization in the midline. These trials were excluded from all further analyses.

3. Procedure

The procedure was the same as in experiment 1a except that the to-be-adjusted variable during the compensation phase was the ITD.

4. Design

The independent variable was the IID, which was manipulated within subject in seven steps (−7.5, −5.0, −2.5, 0, 2.5, 5.0, and 7.5 dB). The dependent variables were (a) the chosen ITD during the compensation phase and (b) the perceived location during the localization phase of the experiment. As mentioned above, we expected a smaller effect size than in experiment 1a because during the compensation phase we tried to shift participants' attention toward the interaural cue that already is the dominant cue. We presumed an effect size of $\eta^2=0.45$. A statistical power analysis using *G*Power3 (Faul *et al.*, 2007) showed that in order to detect effects of $\eta^2=0.45$ of the independent variable given $\alpha=\beta=0.05$, $N=12$ participants were needed.

B. Results

Two participants had to be excluded because for one or more given IIDs they chose the not enough checkbox during all trials. For the remaining participants this was the case for 2.0% of the trials. The left panel of Figure 5 illustrates the relation between the preset IID and the ITD chosen to compensate for the effects of the IID during the compensation

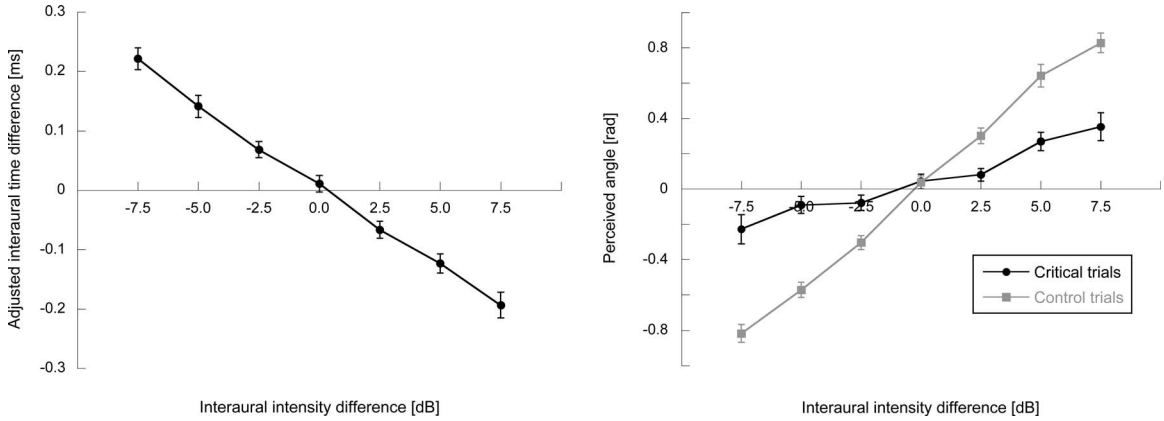


FIG. 5. Left panel: ITDs adjusted by the participants in order to compensate for preset IIDs during the compensation phase of experiment 2 (error bars denote the standard errors of the means). Right panel: Relation between the preset IID and the perceived location during the critical trials and the control trials of the localization phase of experiment 2 (error bars denote the standard errors of the means).

phase. A MANOVA showed that the effect of the preset IID (−7.5, −5.0, −2.5, 0, 2.5, 5.0, and 7.5 dB) on the chosen ITD was statistically significant [$F(6,31)=30.87$, $p<0.001$, $\eta^2=0.86$]. An analysis of the polynomial contrasts showed that only the linear component was statistically significant [$F(1,36)=159.66$, $p<0.001$, $\eta^2=0.82$].

Figure 5 (right panel) shows the relation between the IID and the perceived location during the localization phase. A MANOVA for the control trials showed a significant effect of the IID on the perceived sound source location [$F(6,31)=48.45$, $p<0.001$, $\eta^2=0.90$].

For the critical trials participants had compensated the IIDs by appropriate ITDs during the compensation phase. Thus, one should expect a perceived location at the midline (i.e., a perceived angle of zero) if the processes operating during the compensation and localization phases were identical. As in experiment 1a, this does not appear to be the case, albeit the shift of the perceived location of the critical trials toward the location predicted by the preset interaural cue (IID) was smaller than in experiment 1a. A MANOVA for the critical trials showed that the effect of the IID was nevertheless statistically significant [$F(6,31)=4.75$, $p=0.002$, $\eta^2=0.48$]. An analysis of the polynomial contrasts showed that only the linear component was statistically significant [$F(1,36)=24.30$, $p<0.001$, $\eta^2=0.40$].

C. Discussion

The analysis of the compensation phase data showed that participants had understood the task of canceling out the preset IIDs by appropriate ITDs. However, the data from the localization phase showed that the processes operating while manipulating ITDs to compensate for IIDs cannot be considered equivalent to the processes operating during sound localization in a normal listening situation. The shift-back effect observed in experiments 1a–1c also occurred in experiment 2. However, the effect was smaller in this experiment; this is consistent with the assumption that low-frequency ITDs play a dominant role in sound localization and, hence, attention was directed to the interaural cue that already is the dominant cue.

VI. GENERAL DISCUSSION

The trading ratios found in our experiments are comparable to those reported by Young and Levine (1977) for 500 Hz tones: In experiment 1a we found a trading ratio of 80.1 μs/dB for a preset ITD of 600 μs (79.4 μs/dB for a preset ITD of 500 μs in Young and Levine, 1977); in experiment 2 we found a trading ratio of 27.7 μs/dB for a preset IID of 7.5 dB (40.4 μs/dB for a preset IID of 8 dB in Young and Levine, 1977).

The results of the current experiments show that trading ratios more strongly favored IID when adjusting IID, and more strongly favored ITD when adjusting ITD, relative to a neutral listening context. This implies that during the compensation phases a greater perceptual weight was applied to the to-be-adjusted cue in relation to the fixed cue. This finding suggests that the task of canceling out the effect of one binaural cue by a compensatory value of the other cue implies a shift of attention toward the to-be-adjusted cue which in turn gives more perceptual weight to this cue. This change in the perceptual weights to ITDs and IIDs due to a shift of attention toward one of the cues holds an explanation for both the fact that trading ratios differ depending on which of the cues is regulated as reported by Young and Levine (1977) and for the shift-back effect found in the current experiments. The question arises if our attention approach can be regarded as an extension of the regression interpretation by Trahiotis and Kappauf (1978). What would this mean? It seems possible that attention to one interaural cue (e.g., IID) is accompanied not only by an increased weight of this cue in relation to the other cue but also by the evocation of an internal reference tone with the sound source position of interest (e.g., a central position). This reference tone might be a "remainder" of the practice trials in our experiments or be the result of all-day listening experience. It is plausible to assume that the IID of a reference tone in a central position would be 0 dB. When adjusting the IID of a tone a comparison process between the IID of this tone and the IID of the reference tone (0 dB) could start and lead to a regression to 0 dB.

In contrast to Young and Levine (1977) in our experiments no reference tone was presented. So the plausibility of

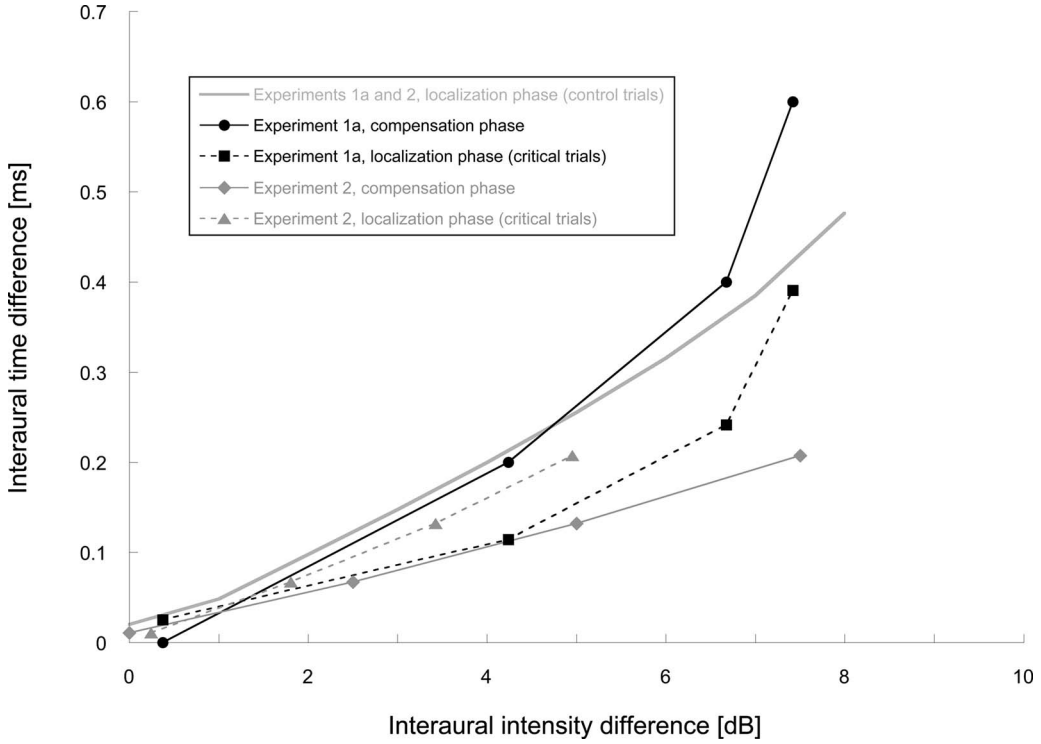


FIG. 6. Equivalence relations during the two phases of experiments 1a and 2 and in the nonconflict situation of the control trials.

the arguments mentioned above depends on the question if participants have internal representations of reference tones that could serve as anchors for regression. In order to test whether an internal representation of a reference tone plays a role in trading experiments in which no (explicit) reference is presented, an experimental design would be needed that prevents the buildup and/or the use of such an internal representation effectively.

Experiments 1b and 1c showed that the shift-back effect can be explained neither in terms of an over-representation of the preset interaural cue (experiment 1b) nor in terms of selective adaptation to the preset and thus constant cue (experiment 1c).

One possible problem has been unmentioned until now. In the localization phases participants had a pointing device which they used to indicate the location of the virtual sound source. This pointing device could have served as a visual reference for the central position. In contrast, during the compensation phases there was no visual reference for centrality. The question arises if this led to a systematic deviation in the position that the participants considered to be the central position in both phases of the experiments. The critical trials with a preset ITD of 0 μs and a preset IID of 0 dB, respectively, can help to determine if such a deviation occurred. In each trial of the compensation phases the starting position of the control slider was set randomly such that the listener had to adjust a compensation value even for the critical trials in which the preset cue was set to zero. The fact that the mean compensation values for a preset cue of 0 μs or 0 dB were very close to zero indicates that there was no systematic error as to the central position during the compensation phases (see Figs. 1, 2, 4, and 5—left panels).

The attentional manipulation should be considered short lived because during the localization phase the sounds were perceived as originating from an off-midline location consistent in its trend with the preset interaural cue. The short-term nature of the attention allocation poses a possible explanation as to why Jeffress and McFadden (1971) found no effect of their manipulation by selective feedback solely on the basis of ITDs or solely on the basis of IIDs. They compared the trials *following* the trials with selective feedback, rather than the trials in which the feedback was given.

In more general terms, the present results clearly show that equivalence relations of ITDs and IIDs depend in part on states of the observer. Thus, the method used to obtain equivalence relations must be taken into account when interpreting them. Relations found by setting both binaural cues into opposition must not be compared with relations found in experiments where only one cue was present at a time (such as the control trials in our experiments).

In order to illustrate the difference, approximations to the current equivalence relations were computed for the two phases and trial types of experiments 1a and 2. In the compensation phases of both experiments and in the critical trials of the localization phases ITDs and IIDs were set into opposition, while in the control trials of the localization phases only one cue was present. Separate equivalence relations were computed for

(a) the control trials of the localization phases of both experiments (no opposition),
(b) the compensation phases of both experiments (the cues were in opposition and attention was directed on one of the cues by the compensation task), and

(c) the critical trials of the localization phases of both experiments (the cues were in opposition but attention was not manipulated).

The equivalence relation of the control trials was obtained as follows. In experiment 1a the data points of the control trials represent the relation between the preset ITD and the perceived location. In experiment 2 the data points of the control trials represent the relation between the preset IID and the perceived location. In order to obtain a function that relates the ITDs to the IIDs, the data points of the control trials of both experiments were approximated by polynomial functions. These functions were set to be equal and resolved for the ITD variable. For both experiments polynomials of orders 1–3 were evaluated in order to determine the best approximation. The best fit to the data points of the control trials of experiment 1a was observed with a third order polynomial [$F(3,3)=6311.4$, $p<0.001$, $R^2=1.000$]. The best fit to the data points of the control trials of experiment 2 was observed with a linear equation [$F(1,5)=1516.5$, $p<0.001$, $R^2=0.997$]. ITDs corresponding to IIDs between −8 and 8 dB in steps of 1 dB were computed. The equivalence relations of the compensation phases of both experiments were obtained by plotting the to-be-adjusted dimension against the preset dimension. The equivalence relations of the critical trials of the localization phases of both experiments were computed as follows. For each preset ITD or IID the perceived location of the *critical trials* was inserted into the regression equation (see above) of the *control trials* of the same experiment. This equation was solved for the corresponding ITD or IID. In this way the ITD or IID was obtained that leads to the same perceived location when presented without a binaural cue of inverse sign. This may be regarded as the residual ITD or IID that had not been compensated for in the compensation phase and thus accounted for the shift-back effect. Hence, the difference between the *preset* ITD or IID and this residual represents the portion of the preset interaural cue that had been compensated for by the adjustment of the other cue.

Figure 6 shows the resulting relations. Comparing the two phases of experiment 1a clearly shows a shift of the equivalence relation toward IIDs in the compensation phase in comparison with the localization phase. A converse shift toward ITDs can be seen in the compensation phase of experiment 2.

It is worth mentioning that most data points are below the curve computed from the control trials. Consistent with the dominant role of ITDs in sounds with low-frequency content, when ITDs and IIDs are set into opposition a shift toward ITDs takes place so that ITDs receive more weight than in a nonconflict situation.

Returning to the central focus of this series of experiments, we want to mention that we currently consider two mechanisms that may operate as a result of the attention shift induced by the task of compensating one interaural localization cue by adjusting the other cue. First, the perceptual weight of the to-be-adjusted cue *per se* may vary along a continuum. Second, the probability of listening to a time image or a time/intensity image may vary.

Whether one of these mechanisms, or perhaps a combination thereof, operates as a result of the attention shifts described here is currently an open question, which may be answered by future research.

[1]The term "trading ratio" suggests a linear relationship between ITDs and IIDs. In the following, the term trading ratio is used either when a linear relationship is assumed or when the relation at a distinct point is reported (e.g., 80.1 μs/dB at an ITD of 600 μs"). In all other cases the more general term "equivalence relation" is used.

[2]In the rest of the article negative ITDs or IIDs denote that a sound was earlier or louder, respectively, on the left channel whereas positive values indicate that it was earlier or louder, respectively, on the right channel.

Algom, D., Adam, R., and Cohen Raz, L. (**1988**). "Binaural summation and lateralization of transients: A combined analysis," J. Acoust. Soc. Am. **84**, 1302–1315.

Breebaart, J., van de Par, S., and Kohlsrausch, A. (**1999**). "The contribution of static and dynamically varying ITDs and IIDs to binaural detection," J. Acoust. Soc. Am. **106**, 979–992.

Colburn, H. S., and Durlach, N. I. (**1965**). "Time-intensity relations in binaural unmasking," J. Acoust. Soc. Am. **38**, 93–103.

Deatherage, B. H., and Hirsh, I. J. (**1959**). "Auditory localization of clicks," J. Acoust. Soc. Am. **31**, 486–492.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (**2007**). "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," Behav. Res. Meth. **39**, 175–191.

Gaik, W. (**1993**). "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," J. Acoust. Soc. Am. **94**, 98–110.

Hafter, E. R., and Carrier, S. C. (**1972**). "Binaural interaction in low-frequency stimuli: The inability to trade time and intensity completely," J. Acoust. Soc. Am. **51**, 1852–1862.

Hafter, E. R., Dye, R. H., Wenzel, E. M., and Knecht, K. (**1990**). "The combination of interaural time and intensity in the lateralization of high-frequency complex signals," J. Acoust. Soc. Am. **87**, 1702–1708.

Hafter, E. R., and Jeffress, L. A. (**1968**). "Two-image lateralization of tones and clicks," J. Acoust. Soc. Am. **44**, 563–569.

Herman, G. E., Warren, L. R., and Wagener, J. W. (**1977**). "Auditory lateralization: Age differences in sensitivity to dichotic time and amplitude cues," J. Gerontol. **32**, 187–191.

Jeffress, L. A., and McFadden, D. (**1971**). "Differences of interaural phase and level in detection and lateralization," J. Acoust. Soc. Am. **49**, 1169–1179.

Leakey, D. M., and Cherry, E. C. (**1957**). "Influence of noise upon the equivalence of intensity differences and small time delays in two-loudspeaker systems," J. Acoust. Soc. Am. **29**, 284–286.

McFadden, D., Jeffress, L., and Ermey, H. L. (**1971**). "Differences of interaural phase and level in detection and lateralization: 250 Hz," J. Acoust. Soc. Am. **50**, 1484–1493.

McFadden, D., Jeffress, L., and Lakey, J. R. (**1972**). "Differences of interaural phase and level in detection and lateralization: 1000 and 2000 Hz," J. Acoust. Soc. Am. **52**, 1197–1206.

McFadden, D., Jeffress, L. A., and Russell, W. E. (**1973**). "Individual differences in sensitivity to interaural differences in time and level," Percept. Mot. Skills **37**, 755–761.

Mills, A. W. (**1972**). "Auditory localization," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. **11**, pp. 303–348.

Phillips, D. P., and Hall, S. E. (**2005**). "Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level," Hear. Res. **202**, 188–199.

Sheldon, P. E. (**1973**). "Equal-onset contours of vibrotactile stimuli," Percept. Psychophys. **13**, 403–407.

Shinn Cunningham, B. G., Santarelli, S., and Kopco, N. (**2000**). "Tori of confusion: Binaural localization cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636.

Trahiotis, C., and Kappauf, W. (**1978**). "Regression interpretation of differences in time-intensity trading ratios obtained in studies of laterality using the method of adjustment," J. Acoust. Soc. Am. **64**, 1041–1047.

Whitworth, R. H., and Jeffress, L. A. (**1961**). "Time vs. intensity in the localization of tones," J. Acoust. Soc. Am. **33**, 925–929.

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. (**1997**). "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 1–23.
Young, L., and Levine, J. (**1977**). "Time-intensity trades revisited," J. Acoust. Soc. Am. **61**, 607–609.

Discrimination and identification of azimuth using spectral shape[a)]

Daniel E. Shub[b)]
Speech and Hearing Bioscience and Technology Program, Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
and Hearing Research Center, Biomedical Engineering Department, Boston University, Boston, Massachusetts 02215

Suzanne P. Carr
Hearing Research Center, Biomedical Engineering Department, Boston University, Boston, Massachusetts 02215

Yunmi Kong
Hearing Research Center, Biomedical Engineering Department, Boston University, Boston, Massachusetts 02215
and Princeton University, Princeton, New Jersey 08544

H. Steven Colburn
Hearing Research Center, Biomedical Engineering Department, Boston University, Boston, Massachusetts 02215
and Speech and Hearing Bioscience and Technology Program, Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 26 February 2007; revised 24 June 2008; accepted 8 August 2008)

Monaural measurements of minimum audible angle (MAA) (discrimination between two locations) and absolute identification (AI) of azimuthal locations in the frontal horizontal plane are reported. All experiments used roving-level fixed-spectral-shape stimuli processed with nonindividualized head-related transfer functions (HRTFs) to simulate the source locations. Listeners were instructed to maximize percent correct, and correct-answer feedback was provided after every trial. Measurements are reported for normal-hearing subjects, who listened with only one ear, and effectively monaural subjects, who had substantial unilateral hearing impairments (i.e., hearing losses greater than 60 dB) and listened with their normal ears. Both populations behaved similarly; the monaural experience of the unilaterally impaired listeners was not beneficial for these monaural localization tasks. Performance in the AI experiments was similar with both 7 and 13 source locations. The average root-mean-squared deviation between the virtual source location and the reported location was 35°, the average slopes of the best fitting line was 0.82, and the average bias was 2°. The best monaural MAAs were less than 5°. The MAAs were consistent with a theoretical analysis of the HRTFs, which suggests that monaural azimuthal discrimination is related to spectral-shape discrimination. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2981634]

PACS number(s): 43.66.Qp, 43.66.Ts, 43.66.Pn [RYL] Pages: 3132–3141

I. INTRODUCTION

Research on azimuthal localization, especially for sources located in the frontal portion of the horizontal plane, has traditionally focused on binaural performance where the most important information is carried by the interaural differences in time delay (ITD) and level (ILD). When interaural comparisons are not possible (e.g., due to a severe unilateral hearing loss), subjects must rely on spectral-shape or overall-level information, which is primarily determined by the head shadow and pinna effects (Shaw, 1974; Shaw and Vaillancourt, 1985). The limits of localization performance based on interaural comparisons are well known [e.g., see Blauert (1997)], but the ability to use the information in the spectral shape is less well characterized. Studies of monaural localization that utilize a roving-level stimulus to reduce the overall-level information include those of Slattery and Middlebrooks (1994) and Wightman and Kistler (1997). More recently, there have also been a number of monaural localization studies with cochlear implant (CI) subjects [e.g., see van Hoesel *et al.* (2002), van Hoesel and Tyler (2003), Litovsky *et al.* (2004), Nopp *et al.* (2004), and Poon (2006)].

The motivation for this study comes from Blauert (1982) where he noted that there is a difference between reporting the perceived location of the source and using all the information to determine the actual location of the source. This distinction could be particularly important for monaural localization in which the normal ITD and ILD cues are missing

[a)]Parts of this work were presented at the 30th Midwinter Meeting of the Association of Research in Otolaryngology.
[b)]Present address: Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104; electronic mail: dshub@sas.upenn.edu

 © 2008 Acoustical Society of America

or in conflict with the spectral-shape cues. In this monaural-listening case, the ITD is undefined and the ILD is consistent with an extreme lateralization toward the presentation side. In fact, most listeners report that a monaural stimulus is perceived at the presentation ear (Slattery and Middlebrooks, 1994; Wightman and Kistler, 1997). In spite of this perception, the spectrum of the received sound provides information about the location of a source. In previous monaural localization studies, subjects were generally requested to "report the apparent location" of the sound as opposed to reporting the location that they believed was correct, and no trial-by-trial feedback was given (Litovsky *et al.*, 2004; Yu *et al.*, 2006; Slattery and Middlebrooks, 1994; Wightman and Kistler, 1997; Lessard *et al.*, 1998). The goal of the current study is to measure monaural localization performance when the subjects are instructed to use all the available information to identify which virtual location was presented and when correct-answer feedback is given.

The monaural stimuli used in these studies are unnatural for normal-hearing (NH) listeners who normally listen with two ears, and one might speculate that subjects with large unilateral losses (ULs) would perform better. When instructed to report the apparent location, chronically unilateral listeners, who normally localize with input to a single ear, performed better than subjects with NH (Slattery and Middlebrooks, 1994) and bilateral CIs (Poon, 2006). The benefits of prolonged monaural experience for monaural localization based on the spectral profile are not known; it is possible that the ability to perform a complex profile analysis is either equally useful for both binaural and monaural listeners or rapidly learned.

The motivation for this experiment was not to learn where subjects perceived monaural stimuli but to see if the monaural spectral profiles associated with azimuthal locations provided subjects with sufficient information to select the correct azimuthal location. This experiment is relevant for the interpretation of performance by bilateral CI users who are asked to identify source directions using a single implant. In interpreting results from studies on the ability of CI users to localize sounds monaurally, the instructions of the experiment may play a confounding role, as in the difference between "relate your impression of where the sound came from" and "try to respond with the correct location." In particular, it is possible that there are sufficient spectral cues for a unilateral CI listener to map out a reliable spectral profile associated with azimuthal locations and "get the right answer" meanwhile perceiving the location of every source as coming from the side of the unilateral implant.

Keeping in mind CI users, the ability of NH listeners to perform monaural localization tasks with constraints similar to those used in CI localization studies was tested. Due to the limited availability of bilateral CI subjects, CI localization studies often do not provide extensive training. Further, CI processing often limits the high-frequency information (up to 8 kHz) and number of channels (approximately 20) available to the listeners. Finally, when listening unilaterally, bilateral CI listeners do not have access to any (even extremely attenuated) interaural difference cues. Therefore, in the current tasks, NH and unilaterally impaired listeners were given little training, and the stimulus was a flat-spectrum tone complex with 20 logarithmically spaced components that had a maximum frequency of 8 kHz. To eliminate the usefulness of all interaural difference cues as well as to provide a standardized set of spectral differences, virtual auditory techniques based on nonindividualized head-related transfer functions (HRTFs) were used. The HRTFs that were used were recorded with an in the ear microphone and not with a behind the ear microphone that is common with CIs. The effects of microphone placement on the HRTFs are likely limited for stimuli with a maximum frequency of 8 kHz because much of the information about the sound source location arises from head shadowing effects and not from pinna filtering.

We utilized both discrimination and absolute identification (AI) of azimuthal-angle paradigms to assess if the spectral profile (or more accurately the aspects of the profile that the subjects utilize) depends on the azimuth in a monotonic manner. The experimental paradigm differs from traditional measurements of localization in four ways. The most obvious difference is that the stimuli were presented monaurally. Two other changes were made to help subjects learn the mapping between potential cues and source locations; the first was that feedback was provided after every trial, and the second was that a fixed reference stimulus, designed to serve as a perceptual anchor (Braida *et al.*, 1984), was presented on every trial. The final change was, as mentioned previously, to use a flat-spectrum tone complex, as opposed to the noise and speech stimuli typical of localization studies. An advantage of using a tone complex, as opposed to a stimulus with a continuous spectrum, is that there is no variability in the short-time spectral shape. Future studies will address the influences of spectral variability (both short time and long time) on monaural localization. With these changes, the current study provides data as to the ability to identify the location of a sound source when all the available information, not just the perceived location, is utilized.

II. EXPERIMENTAL METHODS

Two different experiments were conducted with the same subjects, stimuli, and experimental apparatus. Both experiments used virtual auditory techniques with nonindividualized HRTFs to simulate sound sources with different locations. In the first experiment, the minimum audible angle (MAA) was estimated to characterize listeners' abilities to distinguish between two locations. In the second experiment, AI performance was assessed with a virtual-speaker array (seven or thirteen locations) spanning the frontal hemi-field. In all experiments, correct-answer feedback was provided after each trial.

The MAA task, the 7-location AI task, and the 13-location AI task explore different aspects of monaural localization. The MAA task provides insight into whether or not there is sufficient information in the spectral profile to discriminate between two locations. In the AI tasks, memory requirements are greater with 13 locations than with 7 locations, particularly if there is no percept that varies continuously with azimuth. Thus, similarity in performance [e.g., the root-mean-squared (RMS) difference between the response

location and source location] in these AI tasks could suggest that the subjects use a percept that varies in a systematic manner with source location.

A. Subjects

Six subjects, with ages between 19 and 37 years, participated in both experiments. The subjects had either NH, corresponding to thresholds below 20 dB HL at octave frequencies between 250 and 8000 Hz, or a UL. The UL subjects had normal hearing at one ear and a substantial impairment, thresholds greater than 60 dB HL, at the other ear such that much of their everyday listening was effectively monaural.

The etiologies of the hearing losses varied among the UL subjects. Subject UL1 had unilateral congenital microtia and atresia that resulted in a conductive loss on the left side. Subject UL2 had an idiopathic congenital sensory-neural impairment at her right ear. Subject UL3 had an acoustic neuroma (right side) removed approximately seven years prior to the testing. The stimuli were presented to the normal hearing ear of the UL subjects (the right ear for UL1 and the left ear for UL2 and UL3) and to the right ear of the NH subjects. By including both NH and UL subjects, the benefits, if any, of long-term "monaural" listening on monaural location identification, as measured in this experiment, can be assessed.

The experience of the subjects with psychoacoustic experiments and their familiarity with the questions being explored varied. Subjects NH1, NH2, and UL1 are authors and had prior exposure to the stimuli and paradigms used in this work. In particular, NH1 had extensive exposure to the stimuli and paradigms during pilot listening. Subject UL3 had participated in numerous psychophysical tasks but did not have previous exposure to the stimuli or paradigms used in this work; the subject was also aware of the purposes of the experiment. Subjects NH3 and UL2 were naïve and had never participated in a psychophysical experiment; they received an hourly wage for their participation.

B. Spatial processing

Virtual auditory techniques were used for the spatial processing to eliminate all binaural cues. Virtual locations were simulated by filtering the flat-spectrum tone complex with a HRTF recorded at the right ear of a Knowles Electronics Manikin for Acoustic Research (KEMAR).[1] For all subjects the right-ear HRTF was used (with appropriate ipsilateral and contralateral associations), even when the stimulus was presented to the left ear (UL2 and UL3). The HRTFs of the left and right ears are not identical due to slight asymmetries in KEMAR and the recording apparatus. By only using nonindividualized HRTFs obtained at the right ear, the same stimuli were used for all subjects. Since nonindividualized HRTFs were used in all cases, any effect of using the HRTFs recorded at the right ear of KEMAR when presenting stimuli at the left ear of the subjects should be small. It should also be noted that the use of nonindividualized HRTFs might negate some of the potential benefits of the UL listeners. Specifically, benefits from familiarity with the unique spectral differences associated with their individualized HRTFs are lost. However, benefits from experience using only spectral profiles for localization are preserved.

C. Stimuli

A tone complex with 20 components spaced logarithmically between 500 and 8000 Hz was used in both the MAA and AI tasks. The rise and fall times of the tone complex were 25 ms. The relative phases of each component were randomized on every presentation, and the relative levels of each component (before spatial processing) were identical (i.e., flat spectrum).

The stimuli were band limited to frequencies below 8 kHz for four reasons. First, for frequencies above 8 kHz, the dependence of the spectral profile on location is not monotonic. For example, at some of these higher frequencies, a small change in location toward the side ipsilateral to the recording unexpectedly increases the level. Second, individual differences in HRTFs are more pronounced at frequencies above 8 kHz. Third, future investigations may focus on monaural localization of speech, which has limited energy above 8 kHz. Fourth, CI processors often limit the amount of high-frequency information.

There were two differences in the stimuli used in the MAA and AI tasks. These differences arose because the experiments were designed separately even though they were ultimately run on a common group of subjects. The first difference relates to the duration, and the second relates to the level. The duration of the tone complex in the MAA task was 1000 ms, while in the AI task (both with 7 and 13 locations) the duration was 800 ms. In the MAA task, after the spatial processing, the overall level of the stimulus was roved over a 10 dB range (± 5 dB) around 60 dB SPL on every presentation. In the AI tasks, the overall level, after spatial processing, was first normalized (to reduce the loudness differences across locations) and then roved by ± 4 dB (an 8 dB range).

D. Equipment

During the experiments, subjects sat in a sound-treated room in front of a LCD computer monitor and responded through a graphical interface via a computer mouse. On each trial, "lights" displayed on the monitor denoted the current interval number. The cueing lights were arranged horizontally and therefore moved left to right. Further, in the AI task, a graphical display of the speaker array with the appropriate response locations was continuously displayed to the subjects on the monitor. This graphical display was simply a half circle with small squares representing the possible speaker locations and a small circle representing the position of the listener. The visual angles of the graphical representation did not match the true source locations since the virtual-speaker array spanned the frontal 180° of azimuth, and therefore many of the sources would be out of the angular range of the monitor. In addition, the subject's viewing position of the graphical representation was not controlled. The experiment was self-paced, and listening sessions lasted no more than 2 h with rest breaks. A computer and Tucker-Davis-

Technology System II hardware generated the experimental stimuli at a sampling rate of 50 kHz. Stimuli were presented over Sennheiser HD 265 headphones.

E. Minimum audible angle procedures

In the MAA experiment, subjects were presented a series of two-interval trials. The first interval was presented from the virtual reference location (fixed for each run of 90 trials), and the second interval was from the virtual comparison location (to either the left or right of the reference location). The left and right comparison locations were presented with equal probability, and the subjects were instructed to report whether the comparison location was to the left or to the right of the reference location. A level randomization (chosen independently from interval-to-interval and trial-to-trial) was applied to the stimulus in each interval. There was 200 ms of quiet between each interval. Trials were self-paced, and the subjects received correct-answer feedback after every trial.

The probability of a correct response was measured at three reference locations with four different comparison angles for each reference. For each reference location and comparison angle, 90 trials were conducted. For each reference location, all 360 trials, 90 for each of the four comparison angles, were completed before the next reference location was tested. The ordering of the comparison angles was always 30°, 15°, 10°, and 5°. The ordering of the reference location relative to the midline was the same for all subjects; specifically, the ordering was (1) 60° to the left of the midline, (2) at midline, and (3) 60° to the right of the midline. Therefore, the right-ear subjects started with the reference location on the side contralateral to the presentation ear, and the left-ear subjects (UL2 and UL3) started with the reference location on the side ipsilateral to the presentation ear.

In the analysis of the MAA task, the difference between the reference and comparison angles is used and not the difference between the left and right comparison angles. Although an ideal observer would ignore the reference presentation and decide if the second interval was the left or right comparison, many of the subjects reported basing their decisions on whether the comparison was to the left or right of the reference. This strategy is consistent with attempting to reduce memory demands by not making comparisons to a standard held in memory.

F. Absolute identification procedures

In the AI task, subjects were presented a series of three-interval trials. The first and third intervals cued the subjects to the extremes of the virtual-speaker array. Specifically, the first interval of each three-interval trial was presented from the extreme location (90° away from midline) on the side contralateral to the presentation ear, and the third interval was presented from the extreme location on the side ipsilateral to the presentation ear.[2] The second interval was presented from a random virtual location, and subjects were asked to identify the location of the source on the second interval. The level randomization applied to each interval was chosen independently. There was 800 ms of quiet between each interval. Trials were self-paced, and subjects received feedback after every trial. The feedback informed them of the actual source number both through a textual message and by highlighting the correct location in the graphical representation of the speaker array. The subjects were told that the goal was for their responses to be as close as possible to the correct answers.

Data were collected for both the 7-virtual-speaker array and the 13-virtual-speaker array. With both arrays, the virtual speakers were equally spaced across the frontal hemifield. The separations between locations were 30° and 15° for the 7-location and 13-location arrays, respectively. For the 7-location task, data were collected in runs of 70 trials (ten presentations from each location), while for the 13-location task, data were collected in runs of 65 trials (five presentations from each location). All subjects completed all five runs of the 7-location task prior to the collection of the data for the 13-location task. In this 13-location task, all subjects completed five runs except UL2 who completed only three runs due to limited availability.

Since the first interval was always from the extreme location on the side contralateral to the presentation ear and the third interval was always from the extreme location on the side ipsilateral to the presentation ear, subjects received identical acoustical stimuli regardless of the presentation ear. However, there was an asymmetry between left- and right-ear listening; the virtual locations moved left to right for the right-ear listeners and right to left for the left-ear listeners (UL2 and UL3). Additionally, the interval markers (cueing lights) displayed on the monitor always moved left to right. Therefore, the interval markers and the sounds moved in the same direction for right-ear listeners, while for left-ear listeners, the interval markers and the sounds moved in opposite directions. Subject UL3 (a left-ear subject) spontaneously reported that the dissonance between the lights and the sounds was "annoying." The impact of these subtle differences between left- and right-ear listening is discussed further below.

The data from the AI tasks were collected in terms of source-response confusion matrices; each entry of these matrices contains the frequency of a particular response (i.e., speaker number) being reported conditioned on the location of the presented stimulus. Confusion matrices can be visualized as dot scatter plots where the diameter of each dot is proportional to the frequency of occurrence of the associated source-response pairing. The source-response confusion matrices can also be characterized by a number of statistical measures. The slope and intercept of the best fitting (least-mean-squared difference) straight line as well as the RMS differences and the mean-unsigned (MUS) differences between the response and source locations are calculated. The RMS and MUS differences are calculated for all source locations together as well as separately for source locations ipsilateral and contralateral to the presentation ear. For completeness, the RMS and MUS differences are also calculated for the center location since this location is not included in either the ipsilateral or the contralateral calculations.

G. Familiarization and training

Subjects were given only limited familiarization and training[3] similar in duration to the familiarization and training stages of studies involving CI listeners. Before each task began, the subjects were given a short demonstration of the stimuli. Immediately following the demonstration, subjects were generally given a short training block to familiarize them with the experimental paradigm. Since subjects NH1, NH2, UL1, and UL3 had experience in discrimination procedures, they were not given a training block in the MAA task (they were, however, given a training block before the 7-location and 13-location tasks). Immediately following the training block, data collection for the appropriate task was started. The MAA task was conducted first followed by the 7-location task and, finally, by the 13-location task.

The demonstration consisted of a stimulus being presented from each virtual location (7 locations in the MAA and 7-location tasks and 13 locations in the 13-location task) from right to left and then from left to right (independent of the ear at which the stimulus was being presented). Following this presentation, stimuli from the extreme left location, then from the middle location, and finally from the extreme right location were presented. During all of these presentations, the cartoon of the speaker array described above was displayed on the monitor, and the currently active location was highlighted.

The short training blocks consisted of a few trials to allow the subjects to familiarize themselves with the experimental procedures. In the MAA task, the practice block consisted of 90 trials with the first reference angle and a change in angle of 30°. In the 7-location AI task, the practice block consisted of two trials from each location, while in the 13-location task, the practice block consisted of one trial from each location.

III. ANALYSIS OF THE SPECTRAL PROFILE

The obvious physical cue, for monaural localization of a fixed-spectrum random-level source, is the spectral shape (magnitude as a function of frequency) of the HRTF since it varies significantly with azimuthal angle. Figure 1 shows the spectral shape of the HRTF for the locations used in the seven-location AI task. The levels of the low-frequency components (below 1000 Hz) change by only a few decibels as a source is moved from the extreme left to the extreme right (as would be expected from the minimal head shadow effect at low frequencies). The levels of the middle frequencies (1000–8000 Hz) show strong dependencies on location and generally increase as the source is moved from the side contralateral to the recording (i.e., the right) toward the side ipsilateral to the recording (i.e., the left). Since the stimuli used in this work do not include frequencies above 8000 Hz, the high-frequency regions of the spectral profiles are not shown.

A spectral analysis of the HRTFs was used to calculate the performance of the ideal observer in the MAA task. The analysis is based on the model of Durlach *et al.* (1986) in which the activity in separate frequency channels is assumed to be statistically independent and to be combined optimally

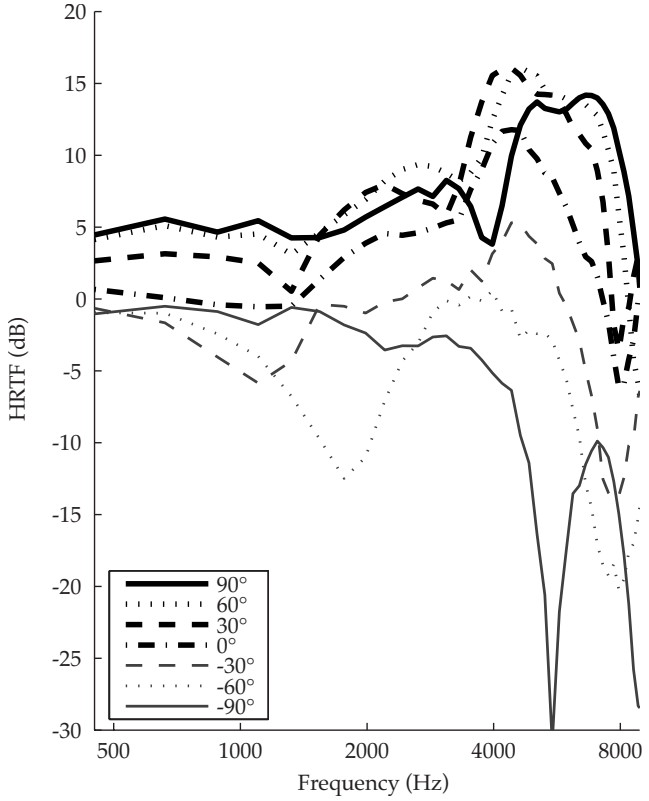


FIG. 1. Spectral profiles as a function of frequency for seven different locations in the frontal portion of the azimuthal plane. The spectral profiles are from right ear HRTF recordings of KEMAR. The same line styles are used for ipsilateral and contralateral locations; however, the linewidth for ipsilateral locations is wider than for contralateral locations. In the psychophysical experiments, the overall differences are reduced by normalization and level randomization, forcing the subjects to rely on differences in the spectral shapes.

such that the square of d' is equal to the sum of the squares of the d' values for the separate channels. The value of the predicted d' then is the product of the square root of the number of independent frequency bands and the RMS difference in the spectral levels for each band divided by the standard deviation of the internal noise. An internal noise with a standard deviation of 1 dB is consistent with thresholds from profile analysis studies involving relatively simple changes to tone complex stimuli [see Green 1988]. Farrar *et al.* (1987) estimated a larger standard deviation of the internal noise (between 4 and 13 dB) for the discrimination between two stimuli with continuous spectra and complicated spectral shapes.

In the current analysis, the standard deviation of the internal noise was chosen (*post hoc*) to be 8 dB, leading to predictions that roughly match the range of performance demonstrated by our subjects in the MAA experiments. In making the predictions we assume that each of the 20 components, which are logarithmically spaced between 500 and 8000 Hz, falls, into independent frequency bands. For comparison with the MAA results, d' is converted into the probability of a correct response by evaluating the standard normal cumulative probability distribution at one-half the predicted d'. The ideal observer ignores the reference and makes a decision based only on the comparison angle. The predictions do not take into account that many of the subjects

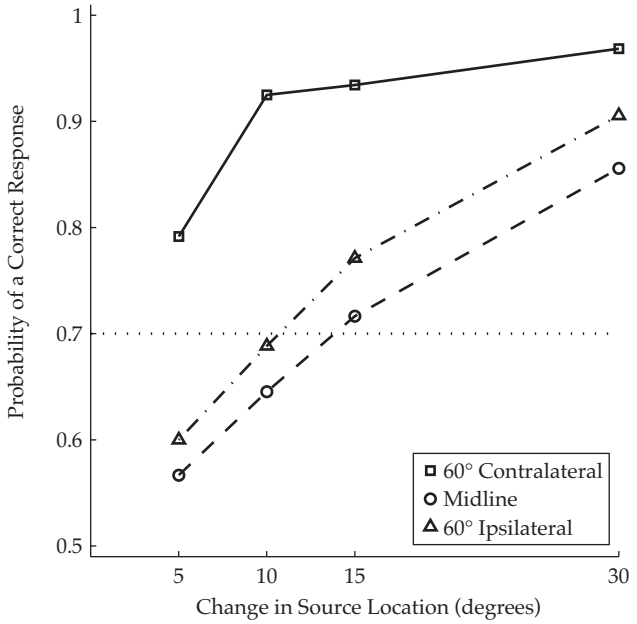


FIG. 2. Psychometric functions in the monaural MAA task for the ideal observer with an internal noise with a standard deviation of 8 dB. The solid, dashed, and dash-dotted psychometric functions correspond to source locations of 60° contralateral (square) to the presentation side, midline (circle), and 60° ipsilateral to the presentation side (triangle). The change in source location is the difference between the reference and left/right comparison locations, not the difference between the left and right comparison locations. The dotted line at 0.7 is a commonly used reference for threshold performance.

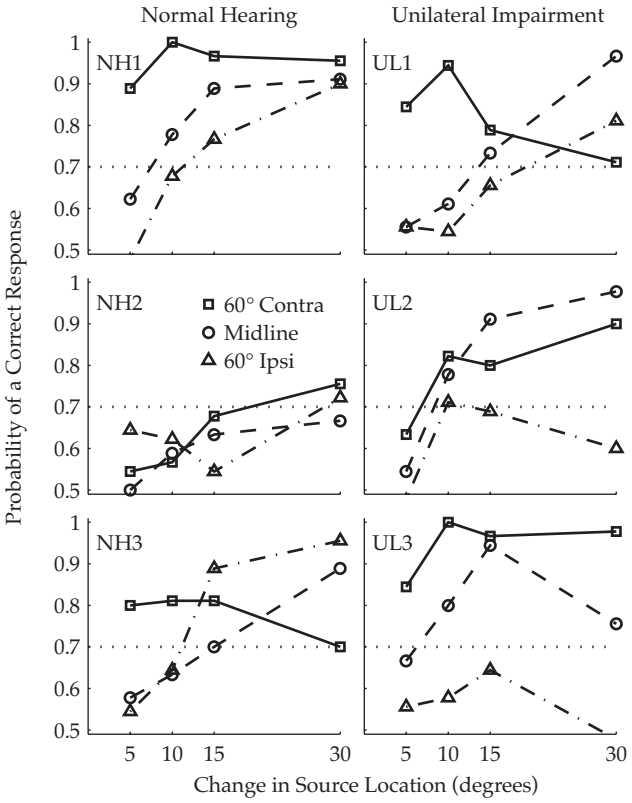


FIG. 3. Psychometric functions for the monaural MAA task for the six subjects. The panels in the left column are for the NH subjects, and the panels in the right column are for the unilateral-impairment subjects. The solid, dashed, and dash-dotted psychometric functions correspond to source locations of 60° contralateral (square) to the presentation side, midline (circle), and 60° ipsilateral to the presentation side (triangle). The presentation side was on the right for all subjects except UL2 and UL3. As in Fig. 2, the change in source location is the difference between the reference and left/right comparison locations, and the dotted line is a commonly used reference for threshold performance.

reported comparing the reference and comparison angles in the MAA task.

Figure 2 shows the predicted MAA psychometric functions (probability of a correct response as a function of the magnitude of the difference between the reference and comparison angles). As one might expect, the probability of a correct response generally increases with increasing spatial separation. Further, performance is greater for the 60° contralateral reference location than for the 60° ipsilateral reference location and is worst for the midline reference location. The probability of a correct response ranges from near chance with the smallest spatial separations at the midline reference location to better than 0.95 with the largest spatial separations at the 60° contralateral reference location.

IV. RESULTS

A. MAA results

Figure 3 shows the MAA psychometric functions obtained for the six subjects. Some of the psychometric functions are nonmonotonic even though the modeling suggests that increasing the spatial separation increases the information (see Fig. 2). Every subject, except NH2, was able to discriminate left from right for some of the tested reference angles and spatial separations. All subjects, except for NH2, were able to identify reliably (probability of correct greater than 0.7) if the comparison angle was 15° to the left or right of the reference angle for the 60° contralateral reference location (squares) and the midline location (circles). For large increments, subjects NH1, NH3, and UL1 could reliably discriminate left from right for the 60° ipsilateral reference location (triangles), while NH2, UL2, and UL3 could not. From these MAA data, it is notable that the two populations (NH listeners and unilateral listeners) are overlapping in their abilities to perform these MAA tests with minimal training.

Learning effects are the likely cause of two aspects of the data. The non-monotonic shape of some of the measured psychometric functions may be an effect of learning because for each reference angle, all subjects were tested with the largest difference between the reference and comparison angles first and the smallest difference last. Learning effects may also be the reason that subjects UL2 and UL3 (the only left-ear subjects) were unable to achieve threshold performance with the 60° ipsilateral reference location that was tested first, but two of the three right-ear subjects (all except NH2), who were tested in the ipsilateral condition last, were able to obtain threshold performance. These effects require more testing on a larger group of subjects to be confident of the factors that are at play and to separate intersubject differences from learning effects.

Apart from subject NH2, the overall level of performance is roughly consistent with the analysis of the spectral profile (see Fig. 2) when the internal noise is assumed to have a standard deviation of 8 dB. One notable exception is

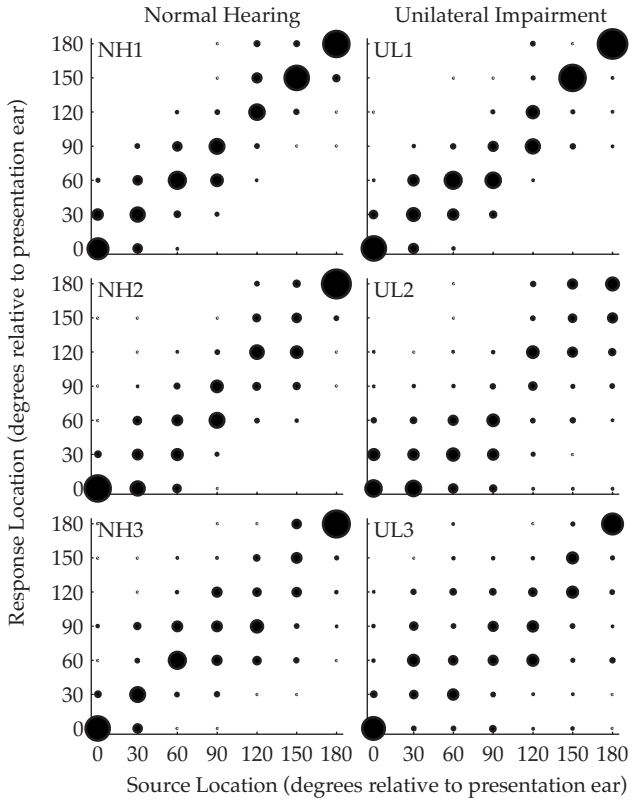


FIG. 4. Dot scatter plots for the AI task with the 7-virtual-speaker array. The diameter of each dot is proportional to the frequency of the source/response pairing. The left column is for the NH subjects, and the right column is for the unilateral-impairment subjects. The source and response locations are relative to the ear at which the stimuli were presented (the right ear for all subjects except UL2 and UL3).

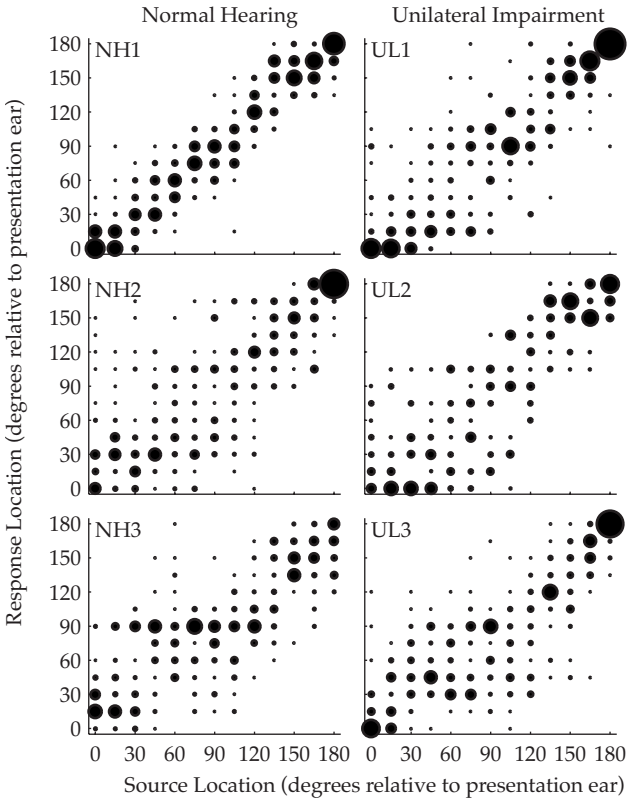


FIG. 5. Same as in Fig. 4, except for the 13-virtual-speaker array.

that the ideal observer performs worst with the midline reference location, while the subjects perform worst with the ipsilateral reference location. The relatively large amount of internal noise in the ideal receiver model needed to approximate the performance of our subjects may be due to the lack of training. It is possible that well trained subjects could perform substantially better on these discrimination tasks.

B. AI results

Figures 4 and 5 show scatter plots for the 7-location and 13-location AI tasks, respectively. As was the case in the MAA task, there were no consistent differences between the subjects with NH and those with unilateral hearing impairments. Overall, the performance in the AI task is quite good. For all subjects, the larger dots in the scatter plots generally lie near the major diagonal. Only a limited number of response locations differed from the source location by over 30°. Visually it seems clear that subjects are able to identify the location of a source within about 30°, consistent with the statistical results shown in Table I. For all the subjects in both tasks, the slopes of the best fitting lines are near unity. The maximum value of the slope is 0.97 and the minimum value is 0.61. The bias (intercept) is generally small, as indicated by the fact that all of the values of the intercepts were within one virtual speaker separation and the majority of the values (7 of 12) were within a quarter of a speaker separation. The measured value of the RMS differences ranged between 25.8° and 50.4° in the 7-location task and 19.1° and 39.3° in the 13-location task. The RMS difference weighs "outliers" (large mistakes) more heavily than the MUS difference. The MUS differences, however, had a similar range (16.0°–34.8°), suggesting that the effects of large errors on performance were small for both the 7- and 13-location tasks.

As in the MAA results, the data are consistent with the existence of some learning between the 7-location and 13-location tasks. Performance, as assessed by the slope, bias, RMS difference, and MUS difference, tends to be better in the 13-location task (tested second) relative to the 7-location task (tested first). In comparing these tasks, most slopes increased toward unity, the magnitudes of the biases decreased in every case, the RMS differences decreased on average and individually in two-thirds of the six cases, and the MUS difference decreased on average (but in only three of six cases). This trend toward a slight improvement suggests that some learning occurred in the course of testing. Whether this learning is related to the AI of location or to the experimental paradigm is unclear.

In contrast to the MAA case, better performance for sources located on the side contralateral to the presentation ear is not seen in the results for either the 7-location task or the 13-location task (see Table I). In these tasks, performance is similar for sources both ipsilateral and contralateral to the presentation ear. Performance for source locations of 0° and 180° is particularly interesting since those locations correspond to the ends of the range of stimuli and are the locations of the reference stimuli that were presented in the first (0°) and last (180°) intervals of each trial. The best perfor-

TABLE I. Summary statistics for the 7-location and 13-location tasks.

			Normal hearing			Unilateral impairment		
			NH1	NH2	NH3	UL1	UL2	UL3
7 locations	Slope		0.91	0.91	0.78	0.90	0.72	0.61
	Bias (deg)		−2.31	6.77	2.22	4.97	−16.5	−8.74
	RMS difference (deg)	All locations	25.8	32.4	37.7	27.5	47.9	50.4
		Ipsilateral locations	26.3	33.4	37.2	26.5	38.3	45.5
		Center location	30.3	31.7	37.0	34.7	53.8	47.4
		Contralateral locations	23.6	31.7	38.5	25.7	54.1	55.7
	MUS difference (deg)	All locations	16.0	21.7	24.2	17.0	34.7	34.8
		Ipsilateral locations	17.4	22.0	21.0	16.6	27.6	31.8
		Center location	21.0	24.0	28.8	28.2	46.2	36.6
		Contralateral locations	13.0	20.6	25.8	13.6	38.0	37.2
13 locations	Slope		0.97	0.77	0.69	0.91	0.89	0.76
	Bias (deg)		0.55	−0.09	0.32	4.06	−4.92	−8.2
	RMS difference (deg)	All locations	19.1	39.3	35.6	32.0	35.0	36.4
		Ipsilateral locations	18.5	44.4	38.5	35.1	38.0	31.8
		Center location	20.6	43.3	32.6	37.2	44.2	37.5
		Contralateral locations	19.4	32.6	33.0	27.4	29.7	40.4
	MUS difference (deg)	All locations	13.6	28.2	27.2	22.5	26.3	25.8
		Ipsilateral locations	13.3	32.5	29.7	26.1	29.7	23.5
		Center location	15.0	38.4	22.8	28.8	34.0	26.4
		Contralateral locations	13.6	22.1	25.4	17.9	21.7	28.1

mance is consistently seen at these two extreme locations. The cueing intervals may have contributed to the similarity in performance for source locations ipsilateral and contralateral to the presentation ear.

V. DISCUSSION

The MAA results show that spectral-shape cues provide sufficient information for the discrimination of angles separated by 10° for many reference directions. This performance is consistent with that expected from spectral-shape discrimination as measured in previous studies (Farrar *et al.*, 1987). The AI results indicate that when feedback is provided so that subjects can use all available information to identify the virtual location, monaural localization performance can be quite reliable with RMS errors around 30°. In addition, it is notable that there were no consistent differences seen between chronically unilateral listeners and NH (bilateral) listeners, even though one might expect an advantage in these monaural experiments for listeners who never have interaural difference cues.

Previous studies found that the perceived location is not very reliable for location judgments with monaural stimuli (Litovsky *et al.*, 2004; Yu *et al.*, 2006; Slattery and Middlebrooks, 1994; Wightman and Kistler, 1997; Lessard *et al.*, 1998). These previous studies generally focused on the perceived location as opposed to subjects' judgments about where the source is actually located and therefore did not provide feedback. The current study employed a paradigm, including feedback, intended to encourage subjects to use any useful cue to optimize the identification of the location of a fixed-spectrum source. We consider the differences between our results and previous monaural localization experiments to reflect this distinction and the fact that the experiments with and without feedback (and with different instructions) measure different abilities.

This distinction between experiments with and without feedback is particularly important when the stimulus is monaural and is naturally perceived at the side of the stimulated ear. Providing feedback allows the subjects to use any information to identify the location of the sound source. Interestingly, the similarity in performance between the 7-location and 13-location AI tasks suggests that the subjects rely on a sound quality that varies systematically with source location. In fact, many of the subjects described attending to the timbre or sound quality of the stimulus for identifying the location of the sound source. Specifically, they reported that stimuli presented from locations contralateral to the presentation ear sounded more "muffled" than stimuli presented from source locations ipsilateral to the presentation ear.

Multiple factors influence performance so that comparisons between the current results and previous monaural localization studies must be done carefully. One clear difference is that unlike in Slattery and Middlebrooks (1994), the subjects in this study do not exhibit a strong bias (i.e., the intercepts of the best fitting lines are near zero). Another difference between the current results and previous monaural localization studies relates to differences in performance for

sources ipsilateral and contralateral to the presentation ear. Slattery and Middlebrooks (1994) reported that performance was better for ipsilateral locations. Wightman and Kistler (1997) argued that the bias could account for the better performance on the ipsilateral side. Further, performance calculated from the optimum use of the spectral profile (see Fig. 2) suggests that performance should be better for contralateral locations. In the MAA task, performance was better for contralateral locations, but in the AI task, there were no substantial differences in performance for sources ipsilateral and contralateral to the presentation ear.

Another notable difference between the current results and previous monaural localization studies is the lack of an advantage associated with prolonged monaural experience (Poon, 2006; Slattery and Middlebrooks, 1994). The lack of a difference in performance between the NH and unilateral-impairment populations in this study is somewhat surprising. One possibility is that the ability to perform a complex profile analysis (not necessarily the analysis done by the unilaterally impaired in real world situations) is either equally useful (or not useful) for both binaural and monaural listeners or rapidly learned. It may also be that the benefits of the chosen paradigm (instructions, familiarization, cueing, feedback, and stimuli) are greater for normal-hearing subjects. Additionally, the use of nonindividualized HRTFs in a highly controlled laboratory task (e.g., fixed spectrum and virtual anechoic space) may have negated any advantages that the unilateral-impairment subjects have from prolonged monaural experience. The unilateral-impairment subjects may perform better then NH subjects in real world localization (familiar sounds, head movement, and visual input). Two of the subjects with unilateral impairments (UL1 and UL3), in fact, reported that they believed that the task would be easier with natural speech.

Understanding monaural localization abilities in more realistic situations (e.g., random-spectrum stimuli and reverberation) is particularly important for CI users. For CI listeners, the reported locations of fixed-spectrum sources in anechoic space is poorly correlated with the actual location (van Hoesel *et al.*, 2002; van Hoesel and Tyler 2003; Litovsky *et al.*, 2004; Nopp *et al.* 2004; Poon, 2006). In these studies, the improvement from near-chance performance with one implant to substantially better than chance performance with two implants is interpreted as evidence of a bilateral advantage. With the current paradigm and nonelectric hearing, however, subjects were able to perform reasonably well monaurally, about as well as listeners with bilateral CIs. The localization performance of NH individuals (using only one ear) and CI users (both unilateral and bilateral users) in more realistic environments needs to be measured. The extent to which this difference in monaural localization performance, between electric and nonelectric hearing, is a result of the CI processing (or impairments to the auditory system of CI listeners) or differences in the paradigm still needs to be investigated.

It appears that subtle changes in the paradigm can have measurable effects on performance. For example, the two subjects that listened with their left ears (UL2 and UL3) showed the most improvement between the 7-location (where the left-ear subjects performed worse than the right-ear subjects did) and 13-location tasks (where both performed the same). The paradigm for the left-ear listeners (specifically the incongruity between the cueing lights and the source locations) may have been less intuitive and therefore led to worse initial performance. This is consistent with the report by Hafter and Carrier (1970) in which a similar incongruity decreased the binaural release from masking. The concern about subtle changes in paradigm highlights the need for care when discussing monaural localization performance.

VI. SUMMARY

This work measured monaural performance in three localization tasks, all implemented with virtual stimuli and nonindividualized HRTFs. The three tasks were an azimuthal-angle discrimination experiment (MAA), a 7-location azimuthal-angle AI experiment, and a 13-location azimuthal-angle AI experiment. In general, conditions in the experiments were arranged to maximize performance. Most importantly, subjects were provided with correct-answer feedback and requested to try to get the correct answer using any useful cues. In addition, reference stimuli from the extremes of the range were provided in each trial of the identification experiments so that some context was maintained, reducing the memory demands on the subjects. Subjects in all experiments included both NH listeners and listeners with large unilateral impairments. Unlike previous monaural localization studies, which generally focused on the perceived location with no feedback provided, in the current experiments, subjects were asked to report where they thought the source was located rather than its perceived location, and feedback was provided. With this paradigm, the measured monaural AI performance was notably better than the near-chance localization performance reported in some previous experiments. Further, the response bias that is typically observed in such experiments was absent. Finally, advantages from long-term monaural experience were not observed; both the NH and unilateral-impairment listeners performed similarly on these tasks. The current results demonstrate that the paradigm can influence performance significantly, and considerable care is required in the interpretation of the ability to identify the location of a sound source when listening with a single ear.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD Grant Nos. DC00100 (S.P.C., D.E.S., and H.S.C.), DC005775 (H.S.C.), and DC004663 (Core Center) as well as the NSF REU program through Grant No. EEC-0244110 (Y.K.). We would like to thank Nathaniel I. Durlach for helpful discussions and comments as well as three anonymous reviewers for helpful comments on a previous version of this manuscript.

[1]The HRTFs used were obtained from the database of Algazi *et al.* (2001). There is no reason to suspect that results would have been different if KEMAR HRTFs from the Gardner and Martin (1995) database or the data reported in Shaw and Vaillancourt (1985) were used to process the stimuli.
[2]The cueing intervals provide no additional information and therefore do

not affect the ideal observer. The cues, however, can act as perceptual anchors and may greatly aid the subjects (Braida *et al.*, 1984).

[3]Subject NH1 had substantial exposure to the stimuli and tasks during pilot listening, and subjects NH2 and UL1 had some exposure during pilot listening.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics* (Mohonk Mountain House, New Paltz, NY), pp. 99–102.

Blauert, J. (**1982**). "Binaural localization," Scand. Audiol. Suppl. **15**, 7–26.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge).

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**, 722–731.

Durlach, N. I., Braida, L. D., and Ito, Y. (**1986**). "Towards a model for discrimination of broadband signals," J. Acoust. Soc. Am. **80**, 63–72.

Farrar, C. L., Reed, C. M., Ito, Y., Durlach, N. I., Delhorne, L. A., Zurek, P. M., and Braida, L. D. (**1987**). "Spectral-shape discrimination. I. Results from normal-hearing listeners for stationary broadband noises," J. Acoust. Soc. Am. **81**, 1085–1092.

Gardner, W. G., and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, New York).

Hafter, E. R., and Carrier, S. C. (**1970**). "Masking-level differences obtained with a pulsed tonal masker," J. Acoust. Soc. Am. **47**, 1041–1047.

Lessard, N., Par, M., Lepore, F., and Lassonde, M. (**1998**). "Early-blind human subjects localize sound sources better than sighted subjects," Nature (London) **395**, 278–280.

Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., and Yu, G. (**2004**). "Bilateral cochlear implants in adults and children," Arch. Otolaryngol. Head Neck Surg. **130**, 648–655.

Nopp, P., Schleich, P., and D'Haese, P. (**2004**). "Sound localization in bilateral users of med-el combi 40/40+cochlear implants," Ear Hear. **25**, 205–214.

Poon, B. B. (**2006**). "Sound localization and interaural time sensitivity with bilateral cochlear implants," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1861.

Shaw, E. A. G., and Vaillancourt, M. M. (**1985**). "Transformation of sound-pressure level from the free field to the eardrum presented in numerical form," J. Acoust. Soc. Am. **78**, 1120–1123.

Slattery, W. H., III, and Middlebrooks, J. C. (**1994**). "Monaural sound localization: acute versus chronic unilateral impairment," Hear. Res. **75**, 38–46.

van Hoesel, R., Ramsden, R., and Odriscoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user," Ear Hear. **23**, 137–149.

van Hoesel, R. J. M., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.

Wightman, F. L., and Kistler, D. J. (**1997**). "Monaural sound localization revisited," J. Acoust. Soc. Am. **101**, 1050–1063.

Yu, G., Litovsky, R. Y., Garadat, S., Vidal, C., Holmberg, R., and Zeng, F.-G. (**2006**). "Localization and precedence effect in normal hearing listeners using a vocoder cochlear implant simulation," Assoc. Res. Otolaryngol. Abstr. **29**, 660.

Current-level discrimination and spectral profile analysis in multi-channel electrical stimulation

Matthew J. Goupell,[a] Bernhard Laback, and Piotr Majdak
Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria

Wolf-Dieter Baumgartner
ENT-Department, Vienna University Hospital, Währinger Gürtel 18-20, A-1097, Vienna, Austria

(Received 19 November 2007; revised 11 August 2008; accepted 12 August 2008)

In experiment 1, six cochlear-implant (CI) listeners discriminated a stimulation pattern eliciting equal loudness for each electrode from a stimulation pattern in which the stimulation at one or more electrodes was increased (peak) or decreased (notch). Three cochlear locations and three bandwidths were tested, without and with level roving. Listeners could always detect peaks but not always notches. Increasing the bandwidth beyond two electrodes produced no improvement in just-noticeable differences (JNDs). JNDs for the basal location were higher than for the apical and middle locations, although listeners had highly individual tendencies. In experiment 2, listeners discriminated changes in the peak heights and notch depths. JNDs for higher peaks were better while JNDs for deeper notches were worse than for experiment 1. In experiment 3, listeners discriminated the electrode position of peaks or notches. JNDs were approximately one electrode. In experiment 4, the first three experiments were repeated with large amounts of level roving. There was no evidence that CI listeners performed an across-channel comparison in these tasks.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2981638]

PACS number(s): 43.66.Ts, 43.66.Fe, 43.66.Dc, 43.66.Qp [BCM] Pages: 3142–3157

I. INTRODUCTION

The perception of complex spectral patterns is vitally important to basic auditory tasks. Speech recognition, in particular vowel identification, depends partially on the ability to identify spectral peak frequencies of speech sounds. The perception of the timbre of different musical instruments depends on which harmonics are emphasized in the spectrum. The localization of sound sources in a vertical plane depends on spectral peaks and notches introduced by direction-dependent filtering imposed by the head, pinnae, and torso of the listener. In this study, we investigated the ability of cochlear-implant (CI) users to discriminate different spectral patterns.

The discrimination of a flat-spectrum stimulus from a stimulus with a spectral peak or notch is considered a profile-analysis experiment. Profile analysis has been extensively studied for normal-hearing (NH) listeners (see Green, 1988 for an introduction). Profile-analysis experiments sometimes use wideband noise stimuli, but more often use complex tones of various spectral densities. It is generally accepted that profile analysis of broadband stimuli involves some form of spectral-shape discrimination process or across-channel comparison. Other processes that could be used for peak and notch detection are discrimination of overall intensity or discrimination of intensity within a single auditory channel. The most direct method of showing that a listener is performing a spectral profile analysis is to include overall level roving in the experiment. The first profile-analysis experiments were performed by Spiegel *et al.* (1981) and Spiegel and Green (1982). They measured the discrimination of a flat-spectrum stimulus from a stimulus with a spectral peak with large amounts of overall level roving. They found that just-noticeable differences (JNDs) increased by only 6 dB when the overall level was varied over a 60-dB range, thus supporting the idea that listeners were using an across-channel comparison to perform the task. For a two-up, one-down adaptive procedure, which tracks a detection probability level of 70.7%, it has been theoretically shown that JNDs that are less than 23.5% of the roving range cannot be achieved by intensity discrimination within a single channel alone (Green, 1988, Appendix A). This limit has sometimes been called the "level detection limit."

Another method of showing that a listener is performing profile analysis is to compare JNDs for a target band without background, which is intensity discrimination, with JNDs with a background. If the JNDs are better for the stimuli with a background, the listener must have gained an advantage from an across-channel comparison. Studies by Green *et al.* (1984), Green and Mason (1985), and Versfeld and Houtsma (1991) have produced examples of JNDs becoming better when there is a stimulus background. We will call the improvement in discriminating the intensity of a target when a background is added the "background advantage."

Yet another method of showing that a listener is performing profile analysis is to vary the bandwidth of the spectral peak or notch to include very large bandwidths. Heinz and Formby (1999) reported the performance of listeners in a profile-analysis experiment that used a variety of bandwidths for peak and notch detection. Their results showed that as the

[a]Electronic mail: matt.goupell@gmail.com

 © 2008 Acoustical Society of America

bandwidth of the peak or notch increased, performance improved until the target bandwidth reached a certain percentage of the full bandwidth of the stimulus with background. Thus a transition occurs where no across-channel comparison can be made and only overall intensity can be used as a cue. In this study, we focus on using the level detection limit and the background advantage methods to assess that profile analysis has been performed.

Only a few studies related to profile analysis have been done with CI users. Drennan and Pfingst (2006) tested the ability of CI listeners to detect a peak on a single electrode fixed in the center of the electrode array while varying the number of electrodes in the background in terms of the background spacing. They found that performance decreased as the number of electrodes increased, contrary to the effects of acoustic spectral range found in some studies with NH listeners. Therefore, there was no background advantage. However, this study was possibly confounded by presenting stimuli with a different number of components at different current levels to maintain a constant overall loudness. This is because single-electrode current-level sensitivity decreases with decreasing current level (Nelson *et al.*, 1996; Drennan and Pfingst, 2005). No overall level roving was used for these experiments.

Henry and Turner (2003) and Henry *et al.* (2005) tested the spectral-resolution capabilities of CI listeners, which may be related to spectral profile analysis. They measured the just-resolvable spectral peak spacing in broadband noises with spectral ripples, which were either linearly or logarithmically spaced and had a 30-dB peak-to-valley ratio. The listeners' task was to discriminate between two spectra where the peak and valley positions were reversed. The stimuli were presented through the speech processor of the CI. Henry *et al.* (2005) showed that, on average, NH listeners' spectral peak resolution was 4.84 ripples/octave and eight times better than CI listeners' spectral peak resolution of 0.62 ripples/octave. In this experiment, the overall level was roved over a range of 8 dB, which was not sufficient to eliminate within-channel intensity cues.

Even though CI listeners seem to be much worse at spectral profile analysis than NH listeners, they have better intensity discrimination thresholds for stimuli delivered directly to the electrodes (Pfingst *et al.*, 1983; Nelson *et al.*, 1996). However, CI listeners often have smaller dynamic ranges (DRs) than NH listeners, and the better intensity discrimination only partially compensates for the smaller DR. Roughly speaking, CI listeners can discriminate intensity about as well as NH listeners. If the limiting factor in a profile-analysis task were intensity discrimination, one would expect CI listeners to perform approximately as well as NH listeners. However, if the limiting factor is the across-channel comparison, one would expect larger JNDs for CI listeners than for NH listeners. This is because the large spread of electrical excitation from a given electrode would effectively smear the spectral information, reducing the spectral contrast between channels, which would make the profile cue less salient. If sensitivity to spectral peaks and notches in the sound spectrum is particularly poor, this may be the reason why speech understanding is limited in CI listeners, which was postulated by Henry *et al.* (2005), and it would make vertical-plane sound localization difficult for CI listeners.

Our experiment follows the basic outline of a study by Moore *et al.* (1989), which we will refer to as simply Moore *et al.* They measured the sensitivity of NH listeners to spectral peaks and notches in noise stimuli. They tested at two center frequencies (1 and 8 kHz), with three different bandwidths, and performed several types of experiments, three of which we replicated with CI listeners. The first experiment was peak or notch detection, which is the discrimination of a stimulus with a spectral peak or notch from a stimulus with a flat background. The second experiment was peak height or notch depth discrimination, which is the detection of a change in height or depth of an existing peak or notch, respectively. The third experiment was frequency discrimination, which is the detection of the change of the center frequency (f_c) of a peak or notch. A short summary of the findings of Moore *et al.* is that JNDs were higher for notches than peaks. The bandwidth of the peak or notch had a small effect on JNDs for peak or notch detection and peak height or notch depth discrimination, but there was no bandwidth effect for frequency discrimination. The JNDs were generally higher at higher f_c. Overall level roving had relatively little effect on the JNDs.

The analogous experiments that we performed were peak or notch detection, peak height or notch depth discrimination, and electrode discrimination with a background where current levels were adjusted to elicit an equal loudness for each individual electrode. The goal of our experiments was to extend our knowledge of spectral-shape discrimination in electrical hearing and to determine to what extent CI listeners were able to perform across-channel comparisons, especially when level roving was included.

II. EXPERIMENT 1: PEAK AND NOTCH DETECTION

A. Listeners and equipment

There were six listeners in this experiment, each implanted with a MED-EL 12-electrode implant. Table I shows the etiology for each listener. The spacing of the electrodes for this type of implant is 2.4 mm. The numbering of electrodes starts at 1 for the apical-most, low-frequency electrode and ends at 12 for the basal-most, high-frequency electrode. All 12 electrodes were used for all six listeners.

The current scaling is linear for this type of implant. For three listeners, one current unit (cu) corresponded to 7.71 μA for all 12 electrodes. For the other three listeners, some electrodes had different scalings; one cu corresponded to either 4.42 or 13.57 μA.

The stimuli were presented directly to the CI via a research interface box, developed at the University of Innsbruck, Austria. Responses were made via a gamepad attached to a personal computer.

B. Stimuli

The peak or notch detection task required the discrimination of a stimulation pattern eliciting equal loudness for each electrode from a stimulation pattern in which the stimu-

TABLE I. Information about the listeners.

Subject	Etiology	Age (yr)	Duration of deafness (yr)	Implant use (yr)	Hearing aid use (yr)	Implant type
CI14	Progressive	67	11	6	15	C40+
CI16	Progressive	54	8	1	30	Pulsar
CI17	Morbus Meniere	59	1	1.5	14	Pulsar
CI18	Progressive	49	11	2	33	Pulsar
CI23	Progressive	75	8	3	14	C40+
CI24	Progressive	45	5	3	7	C40+

lation at one or more electrodes led to a higher (peak) or lower (notch) loudness than for adjacent electrodes. Figure 1(a) shows a schematic of the stimuli for the peak or notch detection task. The background was presented at a comfortable overall level chosen by the listener. Each electrode was stimulated with a constant pulse rate of 1515 pulses per second (pps). The order of electrode activation corresponded to the clinically used order (1, 7, 2, 8,…, 6, 12), which minimizes channel interactions. The duration of the stimuli was 300 ms and the interleaved stimulation of all 12 electrodes occurred for the entire stimulus duration.

Three bandwidths (BW=1, 2, and 3 electrodes) and three cochlear locations of spectral peaks and notches were used. For BW=3, the apical location used electrodes 4–6, the middle location used electrodes 7–9, and the basal location used electrodes 10–12. For BW=2, the highest electrode in the band of three electrodes was omitted at each location (4–5, 7–8, and 10–11). For BW=1, only the center electrode was used (5, 8, and 11). Conditions were tested with overall level roving absent and present. For the conditions with level roving, the current level for each observation interval was varied over a range of 10 cu (±5 cu) according to a rectangular distribution with an integer step size, the change in cu being constant for all electrodes. In total, there were 36 different conditions [3 BW×3 locations×2 band types (peaks and notches)×2 roving types (absent and present)].

An additional type of stimulus was used after the 36 conditions were run, which is called the "band-only" stimulus. This stimulus had no background; the current levels at the electrodes not in the tested band were set to zero. Hence, the band-only stimuli provided intensity discrimination thresholds. Only level increments (peaks) were used and level roving was absent.

C. Procedure

1. Equal-loudness balancing

The DR was calculated in linear cu and was found for each electrode by manually determining the threshold level and maximum comfortable level. Afterwards, an iterative loudness-balancing experiment was performed so that stimuli were presented at a comfortable level and all 12 electrodes elicited the same loudness.

The loudness balancing consisted of two stages: (1) single-electrode balancing (SEB) and (2) overall-comfortable-level adjustment (OCL). Before the first SEB, a single electrode was adjusted to a comfortable level. This electrode was randomly chosen from electrodes 3 to 10 to avoid the apical- and basal-most electrodes, which are sometimes problematic for CI listeners (e.g., they often have the smallest DRs and are more likely to be deactivated). The current level applied to the electrode began at a random level which was $50\%_{DR} \pm 5$ cu. The listener then adjusted the single-interval presentation to a comfortable level. The adaptation step size started at 4 cu, and was reduced to 2 cu after at least one increase and one decrease in level were made. The listener ended the adjustment after at least two increases and two decreases in level by pressing a button. The final

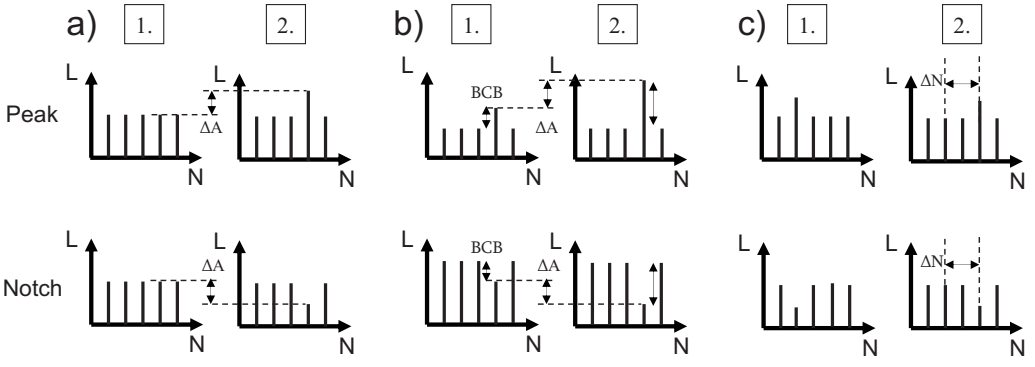


FIG. 1. Schematic of stimuli for the three experiment types: (a) peak and notch detection (experiment 1), (b) peak height and notch depth discrimination (experiment 2), and (c) electrode discrimination of a peak or notch (experiment 3). The plots show the loudness (L) for each electrode (N). An equal-loudness background was used. All three panels show the reference condition for peaks or notches in the first interval. The second interval is the adapted target interval. In experiment 1, illustrated in (a), the height of the peak or depth of the notch, ΔA in cu, was varied. In experiment 2, illustrated in (b), the height of the peak or depth of the notch, ΔA, was varied while keeping the BCB fixed. In both of these experiments, the JND corresponds to average ΔA over several staircases. In experiment 3, illustrated in (c), the distance of the peak or notch between reference and target intervals in electrodes, ΔN, was varied for a fixed peak height or notch depth. The JND corresponds to the average ΔN over several staircases.

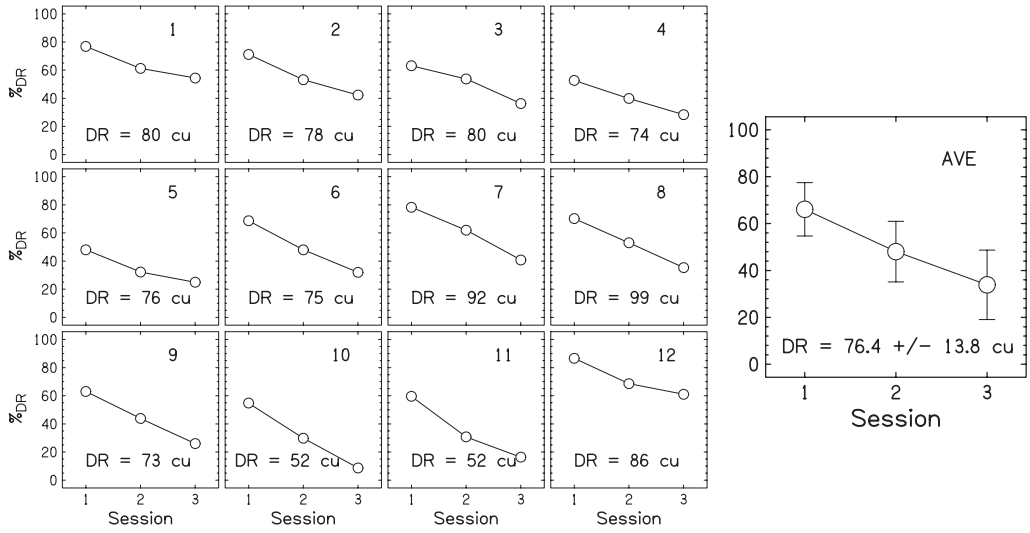


FIG. 2. The results of the loudness-balancing procedure for a typical listener, CI14. The small panels on the left show the results for individual electrodes. Dynamic ranges (DRs) are reported in cu and correspond to $100\%_{DR}$. Data points represent levels that are equally loud between electrodes for a given session. Session 3 shows the data for the final current level used for the equal-loudness stimulus in the main experiment. The large panel on the right shows the averaged $\%_{DR}$ levels over all 12 electrodes. The error bars show ± 1 standard deviation.

value of the level was recorded. After four repetitions of this single-electrode adjustment, the average comfortable level was calculated. The average level from the single-interval presentation was used as a reference for the pairwise level balancings performed in the first SEB.

In the first SEB, 24 balancings were performed (two repetitions for 12 electrodes, presented in random order). The interstimulus interval was 400 ms. In the first interval, the reference electrode was stimulated at the previously found comfortable level. In the second interval, one of the 12 electrodes was stimulated. The level of the second interval started at a random current level that was $50\%_{DR} \pm 5$ cu. The loudness of the second interval was adjusted in 4-cu steps, and this was reduced to 2-cu steps after at least one increase and one decrease in level were made. When the loudness of the second interval matched that of the first interval, after at least two increases and two decreases in level, the listener ended the task by pressing a button. The final value of the level was recorded. The average level of two balancings was calculated for each electrode.

After the first SEB, the first OCL used stimuli that activated all 12 electrodes with interleaved multi-channel stimulation. The current level for the electrodes was the average level found in the first SEB with a change of ± 5 cu, constant across the electrodes. This stimulus was too loud for the listener to be comfortable (by 16 cu per electrode on average). The listener then adjusted the overall level of this stimulus to a comfortable loudness; the step size was 2 cu for all electrodes. The listener ended the adjustment after at least two increases and two decreases in level by pressing a button. Four repetitions of the adjustment were performed and an average level offset was calculated.

Since all the single electrodes were not necessarily at an equal loudness after the first OCL, a second SEB was performed, which was the same as before with one exception: the electrodes (including the reference electrode) were presented at random levels over a 10-cu range centered around the levels determined from the first OCL, not centered at $50\%_{DR}$. The reference electrode was again randomly chosen from electrodes 3 to 10. After the second SEB, a second OCL was performed, which was similar to the first OCL except that it used the levels found in the second SEB. This iterative process continued until the listeners did not change the level of the 12-electrode stimulus in an OCL over the four repetitions (within ± 3 cu on average). Listeners finished the level balancing in two to three blocks of SEB and OCL. The electrode levels determined in the last OCL were used for the main experiment. Figure 2 shows the results of the loudness balancing procedure for an example listener, CI14. From the figure, it can be seen that the curves have different slopes and shapes implying that the loudness growth curves are different for each electrode.

2. Main experiment

A three-interval, two-alternative forced-choice task was used to find the JND for discriminating an equal-loudness background from a background with a peak or notch. The first interval always contained the equal-loudness reference stimulus. The target stimulus with the peak or notch was randomly presented in the second or third interval. The remaining stimulus was the same as the reference. Listeners chose the interval that was different from the other two. The interstimulus interval was 400 ms. Listeners were provided with feedback after each trial.

A two-down, one-up staircase procedure was used to target the 70.7% correct point on the psychometric function. In the staircases, the current level of a band was varied. The level of the band was initially 20 cu different from the reference level for each electrode in the band. Before the first turnaround, the step size was 4 cu, to quickly converge to the approximate JND. After the first turnaround, the step size was reduced to 2 cu. The current level for notch staircases was allowed to go below the single-electrode threshold level,

down to the lowest allowable current level (0 cu). Therefore, turnarounds and notch JNDs below the single-electrode threshold level could occur. If the staircase required a current level below 0 cu, the staircase was terminated with no measurement recorded. Staircases were not allowed to go above the single-electrode maximum comfortable level. The staircase had 16 turnarounds and the average current level in cu of the peak or notch with respect to the background was calculated over the last 12 turnarounds.

For each condition tested, there were at least three staircases. Staircases were monitored for converging characteristics, quantitatively measured by having a within-staircase standard deviation of less than six cu. Learning was also monitored, defined as having average levels monotonically decrease over the three staircases or having the average level of the last staircase be noticeably lower than the others. Additional staircases were performed if there was evidence of nonconvergence within a staircase or suspicion of learning over the staircases. The JND was calculated from the average current level of the last three converging staircases. If the standard deviation of the JND was greater than six cu, more staircases were performed, and the three staircases with the lowest within-staircase standard deviation were chosen for the calculation of the JND.

Before taking data, listeners performed several staircases to learn the task. They were trained until stable performance was obtained. For the actual data collection, the 36 conditions for experiment 1 were presented in random order within a single block. Since there were at least three repetitions for each condition, three full blocks were performed. In cases of nonconvergence, learning, or too large a standard deviation over staircases, additional repetitions were added after the end of the third block. After this, the band-only data were obtained similarly.

D. Results

The JNDs were converted from cu to $\%_{DR}$. For multiple-electrode bandwidths, the DR for the bandwidth was calculated as the average over the single electrodes.[1] The appendix provides JNDs expressed as ten times the logarithm of the Weber fraction and as microamperes.

Figure 3 shows the JNDs for the individual listeners as a function of bandwidth. The height of the small bar, labeled "R," represents the level detection limit for BW=3. Any JND for peaks or notches with roving above this bar may have been achieved by within-channel intensity cues, not by profile analysis. Notice that the listeners have highly individual tendencies, especially with respect to cochlear location. For example, CI14 has relatively small JNDs for the apical and basal locations, but relatively large JNDs for the middle location. CI17 shows the opposite pattern, with the lowest JNDs for the middle location. Hence it is important to plot the individual data.

Figure 4 summarizes the data for all six listeners. The first row of panels shows the JNDs for the peaks averaged over listeners. The second row shows box-and-whisker plots for the notches without roving because JNDs could not be measured for all of the notches and the distribution of measurements was non-Gaussian. Data for notches with roving were omitted because only a few measurements could be made.

It was possible to measure all of the JNDs for the peaks, without and with roving. Averaged over location and bandwidth, the JND for peaks was $7.0 \pm 4.9\%_{DR}$ without roving and $13.5 \pm 6.5\%_{DR}$ with roving. The notches were more difficult to detect than the peaks. Without roving, JNDs could be measured for 69% of all notch conditions (78% apical, 67% middle, and 61% basal; 89% BW=3, 78% BW=2, and 39% BW=1). Averaged over location and bandwidth (for the points that could be measured), the JND for notches without roving was $13.1 \pm 10.1\%_{DR}$. With roving, JNDs could be measured for only 13% of all notch conditions.

Separate three-way repeated-measures analyses of variance (RM ANOVA) for each feature type were conducted (factors: location, bandwidth, and roving). For the data for the peaks, there was a significant effect of location ($p=0.021$). In a *post hoc* test (all *post hoc* tests were Tukey's HSD), there was no difference between the apical and middle locations ($p=0.87$) and between the middle and basal locations ($p=0.090$), but there was a significant difference between the apical and basal locations ($p=0.027$). There was a significant effect of bandwidth ($p<0.0001$). There was a significant difference between BW=1 and the larger bandwidths ($p<0.0001$ for both), but not between BW=2 and 3 ($p=0.096$). There was a significant effect of roving ($p<0.0001$). There were no significant two-way interactions.

A second three-way RM ANOVA was performed on the data for the notches. Missing data for notches were omitted, so the conditions matrix was not full rank. The effect of location was significant ($p=0.001$). JNDs were significantly higher for the basal location than for the apical and middle locations ($p<0.0001$ for both) but there was no difference between the apical and middle locations ($p=0.85$). There was a significant effect of bandwidth ($p<0.0001$), which is easily seen from the box-and-whisker plots in Fig. 4. There was a significant difference between BW=1 and BW=2 ($p=0.001$) or BW=3 ($p=0.006$), but not between BW=2 and 3 ($p=0.84$), as was found for the data for the peaks. The effect of roving was significant ($p=0.002$). The significant two-way interactions were location×roving ($p<0.0001$) and bandwidth×roving ($p=0.015$). Because so few of the JNDs could be measured for the notches with roving, these interactions should be treated cautiously.

The data for the peaks without roving (open circles) can be compared to the band-only data (stars). Qualitatively, the band-only JNDs are smaller than or equal to the peak JNDs. In a three-way RM ANOVA [factors: location, bandwidth, and type (peak or band only)], all three main effects were significant ($p<0.0001$). Thus, there was a significant difference between the peak and band-only data. None of the interactions were significant. A correlation analysis between the peak and band-only data showed a highly significant correlation ($r=0.81$; $p<0.0001$; $n=54$). Focusing on combinations of individual listeners and individual locations (6 listeners×3 locations=18 conditions), there was a signifi-

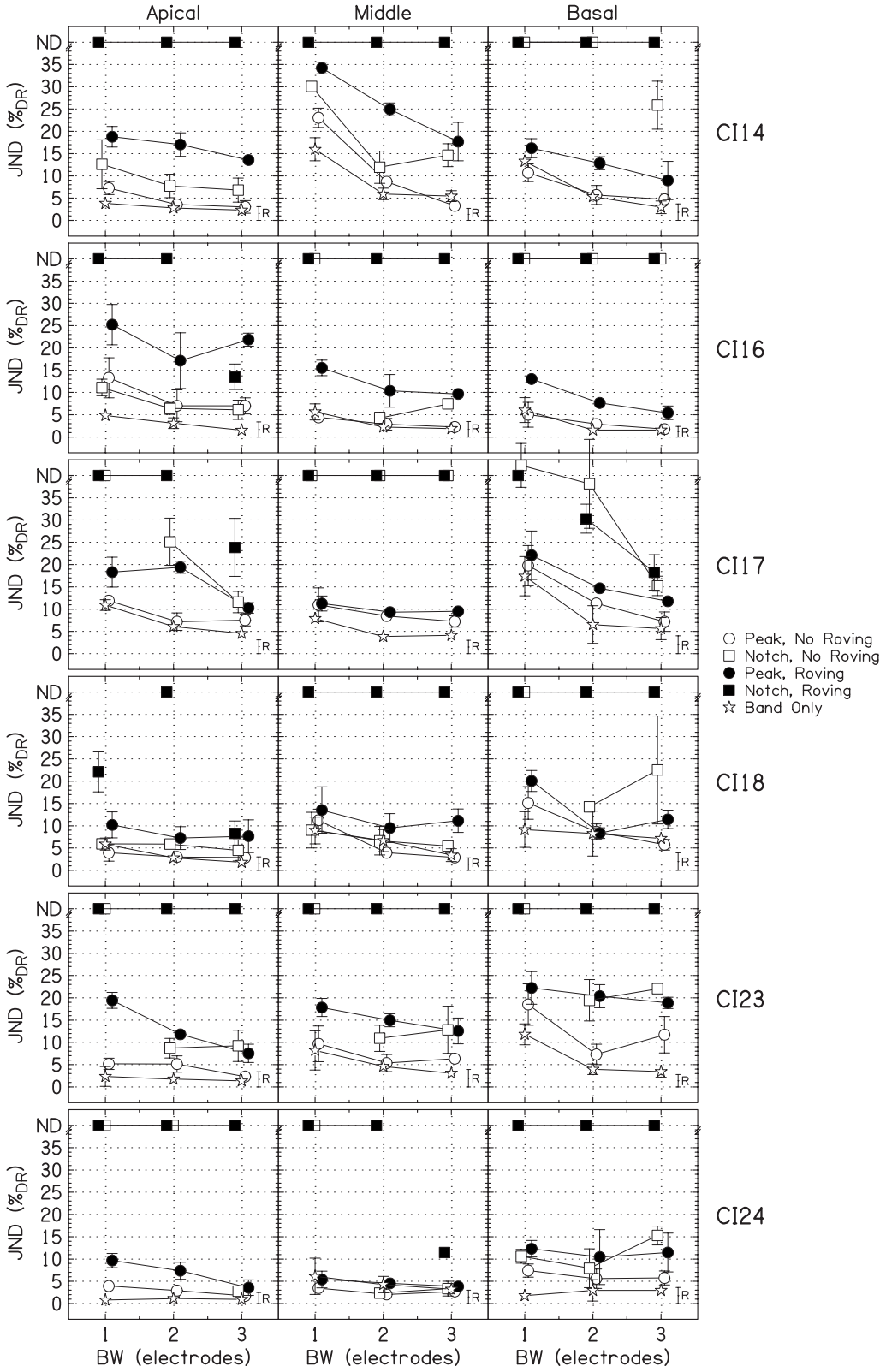


FIG. 3. The results of experiment 1, which show individual JNDs expressed as $\%_{DR}$ for the individual listeners as a function of bandwidth. The circles show the data for the peaks, the squares show the data for the notches, and the stars show the band-only data. The filled symbols indicate the conditions with roving. For notches for which the threshold could not be measured, JNDs are placed at the top of the panels at "ND." The error bars show ±1 standard deviation. The heights of the roving bars, labeled "R," on right side of each panel show the level detection limit for BW=3.

cant difference between the peaks and the band-only data for nine conditions at the 0.05 level, while there was no difference for the other nine.

A separate two-way RM ANOVA was performed on the band-only data. The results were very similar to those for peaks without roving. The effects of location and bandwidth were significant ($p<0.0001$ for both) and the interaction was significant ($p=0.044$). The JNDs for the apical location were

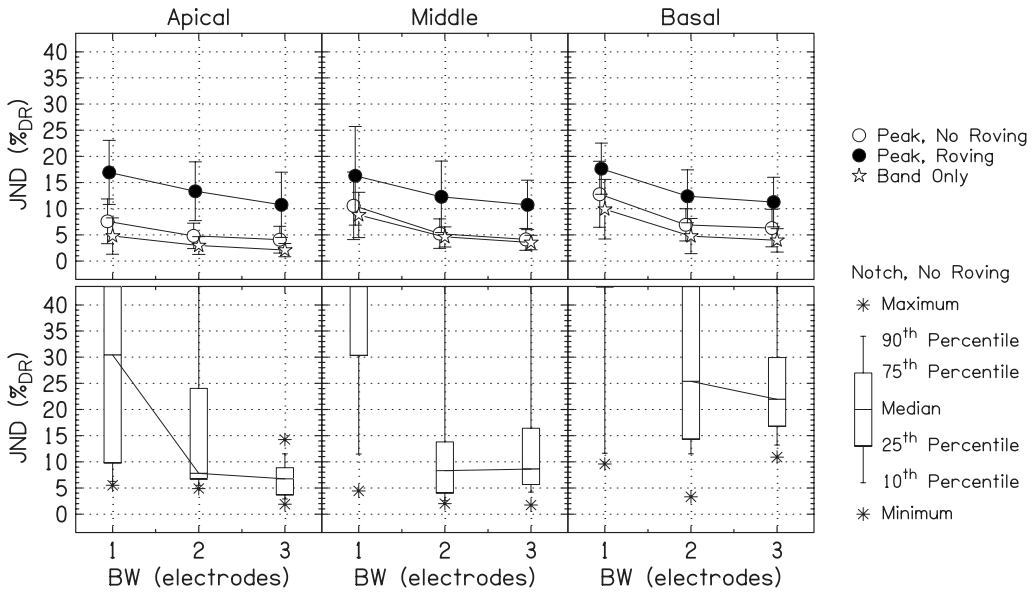


FIG. 4. The average of the JNDs over listeners. The first row of panels shows the average JND for the peaks. The error bars show ±1 standard deviation. The second row shows the box-and-whisker plots for the notches without roving. The percentiles above the top of the plot represent unmeasurable JNDs. The lines connect the medians across bandwidth.

lower than for the middle and basal locations ($p<0.0001$ for both), but there was no difference between the JNDs for the middle and basal locations ($p=0.51$). JNDs for BW=1 were higher than for the larger bandwidths ($p<0.0001$ for both), but there was no difference between the JNDs for BW=2 and BW=3 ($p=0.71$).

E. Discussion

Listeners could always detect sufficiently large spectral peaks, whether or not the level was roved. Notches were more difficult to detect and were sometimes not detectable. A possible reason for the poorer performance in the detection of notches compared to peaks for NH listeners was offered by Moore *et al.* They attributed the poorer notch detection to the spread of excitation from adjacent frequency components: "…a narrow notch in the spectrum of a noise will be represented by only a small dip in the excitation pattern evoked by that noise, even when the notch is infinitely deep." Studies by Ellermeier (1996) and Heinz and Formby (1999) agree with Moore *et al.* The analogous process in CI listeners is the spread of current to the region of the electrode notch from adjacent electrodes. Current from these nearby electrodes effectively fills in the notch. This process would affect peak detection much less.

There was a large amount of interlistener variability in the JNDs for different cochlear locations. This is expected because CI listeners often show highly individualized data in psychophysical tasks, especially with respect to cochlear location. The reason for this is that a number of uncontrollable factors, such as surviving neural populations and placement of the electrode array in the cochlea, can affect the outcome of psychophysical tasks. However, Green and Mason (1985) and Versfeld and Houtsma (1991) also showed individual differences with respect to center frequency in profile-analysis tasks for NH listeners using complex-tone stimuli. Other studies on the effect of center frequency in profile analysis, for example Bernstein and Green (1987), show more uniform results with more extensively trained listeners. Thus, a potential explanation for the individual differences in our data was that the training provided did not give the listeners time to learn the optimal decision strategy for the task. Since we provided training before the experiment, feedback during the experiment (consisting of at least 108 staircases), and we monitored improvements during the experiment, we believe that the individual variability with respect to location of excitation is mostly due to physiological factors and less to procedural factors. This idea is further supported by the intensity discrimination JNDs for the band-only stimuli, which probably required much less training and were highly correlated with the spectral peak JNDs for individual listeners.

There was a significant effect of cochlear location of the peaks and notches, as indicated both by the values of the JNDs and by the number of notch JNDs that could be measured. The largest number of notch JNDs could be measured at the apical location, followed by the middle location, and then the basal location. The NH results of Moore *et al.* showed a corresponding deterioration of JNDs with increasing f_c. One reason for the cochlear location effect for spectral notches in CI listeners could be the equal-loudness stimulus, which tended to have reference levels near threshold level at the basal location. As seen in the example in Fig. 2, electrodes 10 and 11 have equal-loudness currents that are very near threshold. This was typical for four of the six listeners. Thus it appears that our CI listeners tended to have steep loudness growth functions near the threshold level at basal locations.

There was a significant effect of bandwidth, which can be easily seen in Fig. 4. JNDs were significantly higher for BW=1 than for BW=2, but JNDs did not differ for BW=2 and 3. Thus two-electrode peaks and notches are sufficiently

large for the detection of peaks and notches by CI users. A similar trend, especially for notches, was shown for NH listeners by Moore *et al.*

Drennan and Pfingst (2006) performed a peak detection task with Nucleus 24-electrode implant listeners. For the 11-electrode configuration, the peak placed at a single center electrode, and 250 pps per channel (compared to our 1515 pps per channel), the JND, expressed as ten times the logarithm of the Weber fraction, was -7 ± 5 dB. In another experiment using a seven-electrode configuration, the authors found an increase of the JND to −4.9 dB when switching from the 250 pps condition to the conditions most close to our rate (1200–1800 pps). It should be noted that the effect of pulse rate was highly variable across listeners and this approximation represents the average trend. Converting our JNDs for peak detection to the same measure yields -7.9 ± 1.9 dB for BW=3, -7.1 ± 2.3 dB for BW=2, and -4.6 ± 2.6 dB for BW=1. Thus, the JND for the one-electrode condition is similar to that found by Drennan and Pfingst (2006), despite the several differences between the two studies. The most important differences include the stimulation mode (our monopolar versus their bipolar), the number of electrodes in the background, and the level profile (our equal-loudness background versus their background fixed at $70\%_{DR}$).

Drennan and Pfingst (2005) showed no effect of cochlear location on intensity discrimination using electrodes 16 and 8 of the Nucleus24 implant, which correspond to our apical and middle locations, respectively, assuming standard insertion depths. They found an average JND of $4.1\%_{DR}$ for a reference current at $50\%_{DR}$ (monopolar condition). In contrast, for our single-electrode band-only condition, we found significantly higher sensitivity at the apical location ($4.8\%_{DR}$) than at the middle ($8.8\%_{DR}$) and basal ($9.9\%_{DR}$) locations. The worse performance of our subjects at the middle location, which is at odds with the results of Drennan and Pfingst (2005), may be due to a violation of our assumption of standard insertion depth. Our middle location electrodes might be placed more basally than electrode 8 in Drennan and Pfingst's (2005) study.

Our finding that, for half of the conditions tested, JNDs were significantly larger with a background (data for the peaks) than without (band-only data) indicates the relevance of channel interactions, which is consistent with the findings of Drennan and Pfingst (2006). Channel interactions can be viewed as masking from the background electrode currents on the target electrode currents. Moore *et al.* showed for NH listeners that intensity discrimination for a noise band alone worsens when a noise background is added. Hence, both studies showed no background advantage. In contrast, Green and Mason (1985) and Versfeld and Houtsma (1991) showed that profile-analysis JNDs for NH listeners were often better than intensity-discrimination JNDs. The different outcomes may be due to the fact that these studies used complex-tone stimuli whose components fell essentially in separate auditory filters, whereas Moore *et al.* used noise bands with continuous spectra. Therefore, it seems appropriate to consider the stimuli used in this experiment as more comparable to noise bands than to complex tones.

There was a significant effect of level roving, with an increase in the JNDs of approximately half of the roving range on average. This increase is much greater than that shown for NH listeners by Moore *et al.* Also in contrast to our results, many other profile-analysis experiments with NH listeners showed small effects of level roving (Spiegel *et al.*, 1981; Kidd *et al.*, 1989; Drennan and Watson, 2001; Lentz, 2005). The data with level roving can be compared to the bars in Fig. 3, which show the level detection limit for the 10-cu roving range converted to $\%_{DR}$. For most listeners, the JNDs are considerably higher than the bars. Listener CI24 shows JNDs for peaks with roving at the apical and middle locations for the larger bandwidths that are only marginally higher than the level detection limit. Interestingly, many of the JNDs presented in Moore *et al.* are also above the level detection limit. For example, approximately one-third of the peaks for peak detection could have been achieved by using intensity cues within a channel. Therefore, although level roving is extremely detrimental to profile analysis in CI listeners, finding no JNDs below the level detection limit is not necessarily inconsistent with the NH listener results in Moore *et al.*

Because the listeners showed neither a background advantage nor JNDs below the level detection limit, the results raise the concern that the CI listeners did not perform profile analysis, but instead used only intensity cues within a channel. There are several possible explanations for these results, which will be presented in the general discussion. Nevertheless, there remains the possibility that CI listeners can perform profile analysis for stimuli that are not roved in level. When roving is used, it may distract listeners or introduce undesired perceptual changes that cause listeners to perform nonoptimally. NH listeners seem to make nonoptimal use of absolute level when the level is randomized (Kidd *et al.*, 1986, Appendix). The fact that level roving leads to impaired performance does not prove that profile analysis is not performed. Therefore, in the next two experiments, it was useful to assess what performance can be achieved when level is not roved. In experiment 4, all three experiments were rerun with roving ranges of 20 and 30 cu.

III. EXPERIMENT 2: PEAK HEIGHT AND NOTCH DEPTH DISCRIMINATION

A. Methods

In experiment 2, the task of the listener was to identify the interval with the higher peak or deeper notch. The reference intervals had an equal-loudness background with a peak or notch. The height of the peak was either 8 or $16\%_{DR}$. These values were chosen because $8\%_{DR}$ is larger than the average JND for the peak conditions without level roving in experiment 1, thus resulting in a detectable peak for most listeners. The depth of the notches equaled the size of the peaks. These reference peak heights and notch depths will be called the band current with respect to the background (BCB). No level roving was applied.

In the target intervals, the height of the peaks or depth of the notches was varied with respect to the BCB. The peaks and notches had a two-electrode bandwidth and two cochlear

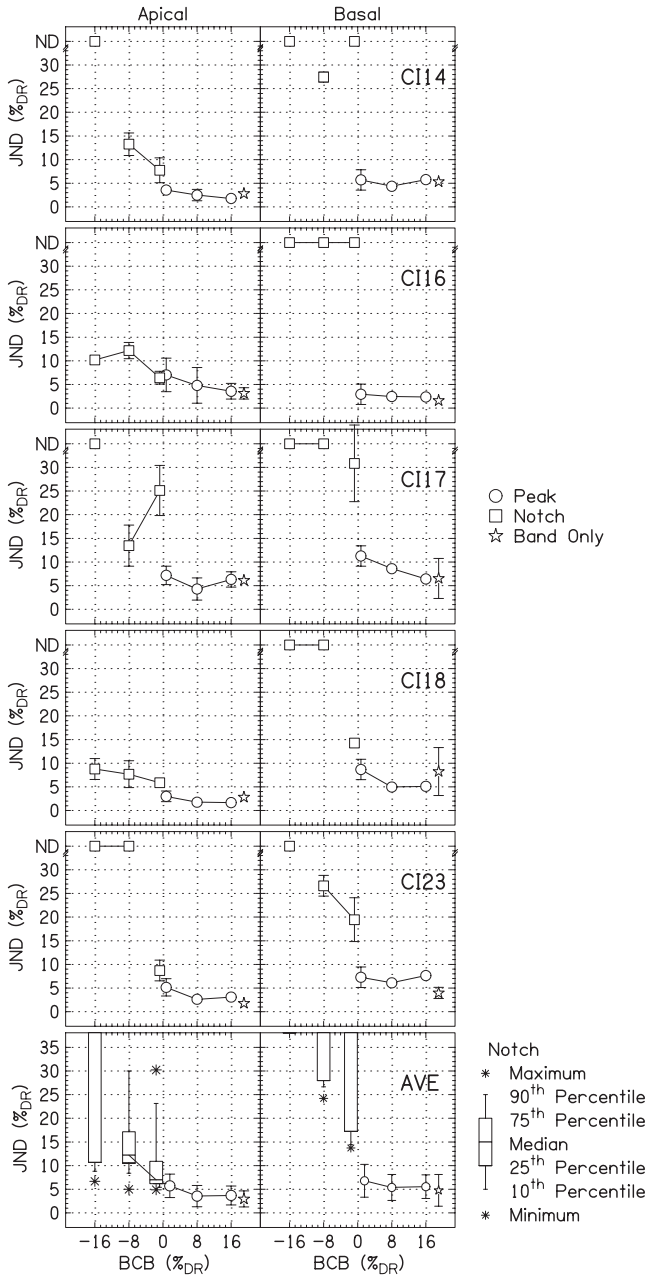


FIG. 5. Individual and average listener data from the peak height or notch depth discrimination task, experiment 2. The JNDs, expressed as $\%_{DR}$, are plotted for different band currents with respect to the background (BCB). The error bars show ± 1 standard deviation. Data for BCB$=0\%_{DR}$ and band-only data are replotted from experiment 1. JNDs for notches that were unmeasurable are placed at the top of the panels at "ND." The notch JNDs are summarized with a box-and-whisker plot.

locations were tested, apical (4–5) and basal (10–11). The two-electrode bandwidth was chosen because beyond that bandwidth no significant improvement in the JND occurred in experiment 1. Figure 1(b) shows an example of a reference stimulus and a target stimulus. The procedure was the same as for experiment 1. The listeners were the same except that listener CI24 did not participate.

B. Results

The results are shown in Fig. 5. Data from experiment 1 were replotted at BCB$=0\%_{DR}$ and were included in the analyses. As for experiment 1, all of the peak JNDs could be measured. The JNDs at the apical location ($3.9 \pm 2.4\%_{DR}$) were significantly lower than those at the basal location ($6.0 \pm 2.8\%_{DR}$), as shown by a two-way RM ANOVA (factors: location and BCB; $p < 0.0001$ for both). There was no interaction between the factors ($p=0.99$). *Post hoc* tests showed that JNDs for BCB$=0\%_{DR}$ were significantly higher than those for BCB$=+8$ and $+16\%_{DR}$ ($p=0.002$ and 0.004, respectively). The JNDs for BCB$=+8\%_{DR}$ were not significantly different from those for BCB$=+16\%_{DR}$ ($p=0.97$).

An additional two-way RM ANOVA combined the data for the peaks with the band-only data from experiment 1 (factors: location and BCB), where the band-only data represent an additional level of factor BCB. The main effects and interactions were similar to those for the previous RM ANOVA. A *post hoc* test showed no difference between results for the band-only condition and the BCB$=+8$ and $+16\%_{DR}$ conditions ($p=0.89$ and $p=0.78$, respectively). Looking at individual pairs of data points, there are three JNDs that are significantly lower than the band-only JNDs as measured by a two-sample t-test: the $+16\%_{DR}$ apical location for CI14 ($p=0.025$), the $+8\%_{DR}$ apical location for CI18 ($p=0.009$), and the $+16\%_{DR}$ apical location for CI18 ($p=0.004$).

As for experiment 1, only a portion of the JNDs for the notches could be measured, 40% of the total. Of these, six could be measured at the apical location and two could be measured at the basal location. In a two-way RM ANOVA for the notches (factors: location and BCB) the effects of location and BCB were significant ($p < 0.0001$ and $p=0.020$, respectively), and the interaction was significant ($p=0.025$). Increasing the depth of the notch made the task more difficult, especially at the basal location.

C. Discussion

The data from experiment 2 are similar in pattern to those from experiment 1 with notches being more difficult to discriminate than peaks. All the JNDs for the peaks were measurable while only 40% of the JNDs for the notches were. As in experiment 1, the basal notches were more difficult to discriminate than the apical notches.

Discrimination of peaks became better with increasing reference peak height and JNDs with BCB greater than $0\%_{DR}$ were not different from those for band-only conditions. This indicates that the detrimental effect of having a background can be compensated by increasing the reference peak height.

Discrimination of notches became worse with increasing notch depth. Reasons for this could be that across-channel interactions effectively reduce the difference in excitation pattern between reference and target stimuli and/or the deeper notches are closer to the electrode threshold level (where no perception is elicited) and 0 cu (where the adaptive procedure terminates). Moore *et al.* also showed that the discrimination of peaks becomes better with increasing peak height and discrimination of notches becomes worse with increasing notch depth for NH listeners.

There is an alternative explanation for the patterns observed in experiment 2. Assuming listeners were not performing profile analysis, but rather using only intensity cues, the task was effectively to detect a change in loudness in which only 2 of 12 electrodes were altered. The overall loudness change therefore depends on the relative loudness of the varied electrodes compared to the remainder of the stimulus. Loudness summation for multiple electrodes has been studied by McKay *et al.* (2001). They found that a louder component (i.e., a higher peak) will dominate the overall loudness and changes in it produce salient overall loudness changes, whereas a softer component (i.e., a deeper notch) contributes negligibly to the overall loudness and changes in it produce nonsalient overall loudness changes. This could explain the better JNDs for higher peaks and worse or unmeasurable JNDs for notches.

IV. EXPERIMENT 3: PEAK AND NOTCH ELECTRODE DISCRIMINATION

A. Methods

In experiment 3, the task was to discriminate the cochlear location of the peak or notch for a fixed peak height or notch depth. An equal-loudness background was present. All three cochlear locations were used as the reference location and the bandwidth was two electrodes. For the apical reference location (4–5), target peaks and notches basal to the reference electrodes were used. For the middle reference location (7–8), target peaks and notches basal and apical to the reference electrodes were used. For the basal reference location (10–11), target peaks and notches apical to the reference electrodes were used. Peak heights and notch depths were always $12\%_{DR}$ averaged over all the electrodes more basal or more apical in location, depending on the target peak or notch location. No level roving was applied.

A schematic of the stimuli is shown in Fig. 1(c). The separation between the electrodes with the peak or notch in the reference and target intervals started at three electrodes. The two-down, one-up staircase procedure decreased the separation by one electrode after two correct answers, and increased the separation by one electrode after an incorrect answer. The location of the peak or notch was bounded by the very edge of the electrode array. If more than four incorrect answers were obtained at the very edge of the electrode array, the staircase was terminated with no measurement recorded. All other procedural aspects were similar to those for experiments 1 and 2. All six listeners participated in this experiment.

B. Results

The results are shown in Fig. 6. The JNDs were measured as electrode separation. For the staircase procedure, the smallest achievable JND is 0.5 electrodes. There were 14 of a total of 48 JNDs that had values very close to 0.5 electrodes. All 24 JNDs for peaks could be measured. However, there were 4 of a total of 24 JNDs for notches that could not be measured, for different listeners, locations, and directions. The average data are arithmetic means calculated over the listeners (the data at "ND" were ignored). The mean JNDs

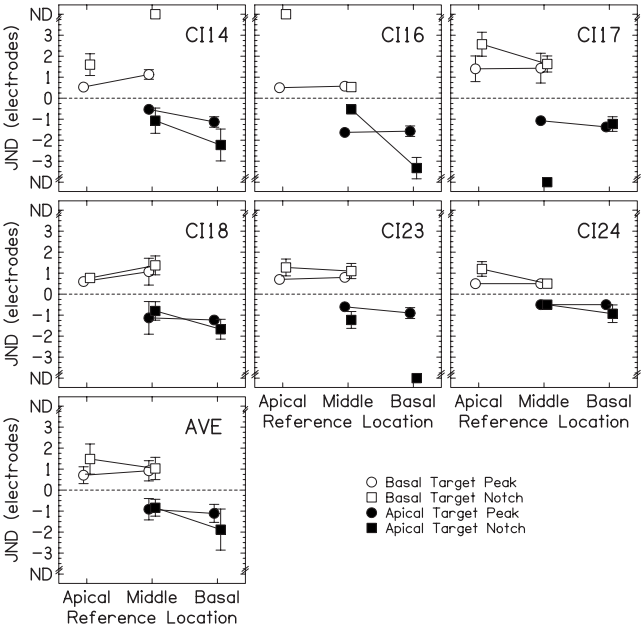


FIG. 6. Individual and average listener data from the electrode discrimination task, experiment 3. Peaks and notches tested at locations basal to the reference location are plotted as positive JNDs and have open symbols. Peaks and notches tested at locations apical to the reference electrodes are plotted as negative JNDs and have filled symbols. The error bars show ±1 standard deviation. JNDs for notches that were unmeasurable are placed at the top and bottom of the panels at "ND."

for peaks and notches were 0.9 ± 0.4 and 1.3 ± 0.8 electrodes, respectively. In terms of tonotopic distance, this translates to 2.2 and 3.1 mm, respectively.

A three-way RM ANOVA [factors: location, direction (from apical and basal), and type (peak and notch)] was performed. There was no effect of direction ($p=0.63$). JNDs for the notches were significantly larger than JNDs for the peaks ($p<0.0001$). There was a significant effect of location ($p<0.0001$). A *post hoc* test showed no difference between the JNDs for the apical and middle locations ($p=0.44$), whereas the JNDs for the basal location were significantly worse than the JNDs for the apical ($p=0.006$) and the middle ($p<0.0001$) locations. The only significant interaction was location × type ($p=0.001$). The peaks and notches led to similar JNDs for the middle location, but notches led to higher JNDs than peaks for the apical and basal locations.

C. Discussion

Listeners could perform the electrode discrimination task for both peaks and notches. Several JNDs were near 0.5 electrodes, the minimum JND measurable in this experiment. It has been shown that pitch percepts intermediate to those elicited by the individual electrodes are possible by varying the level ratio of two neighboring electrodes stimulated simultaneously (Donaldson *et al.*, 2005; Koch *et al.*, 2007; Firszt *et al.*, 2007) or sequentially (McDermott and McKay, 1994; Kwon and van den Honert, 2006). Thus, future studies could measure electrode discrimination for peaks and notches using dual-electrode level adjustment to measure JNDs below the limit of 0.5 electrodes imposed in the present study.

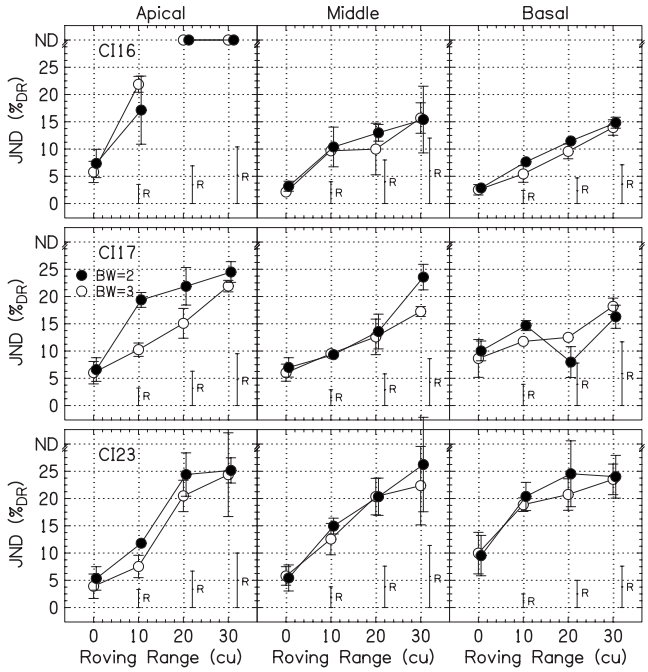


FIG. 7. Individual JNDs for peak detection. Data from experiments 1 (roving range=0 and 10 cu) and 4 (roving range=0, 20, and 30 cu) were combined. The open symbols are for BW=3 and the filled symbols are for BW=2. The height of the roving bars, labeled "R," shows the level detection limit. JNDs that were unmeasurable are placed at the top of the panels at "ND." The error bars show ±1 standard deviation.

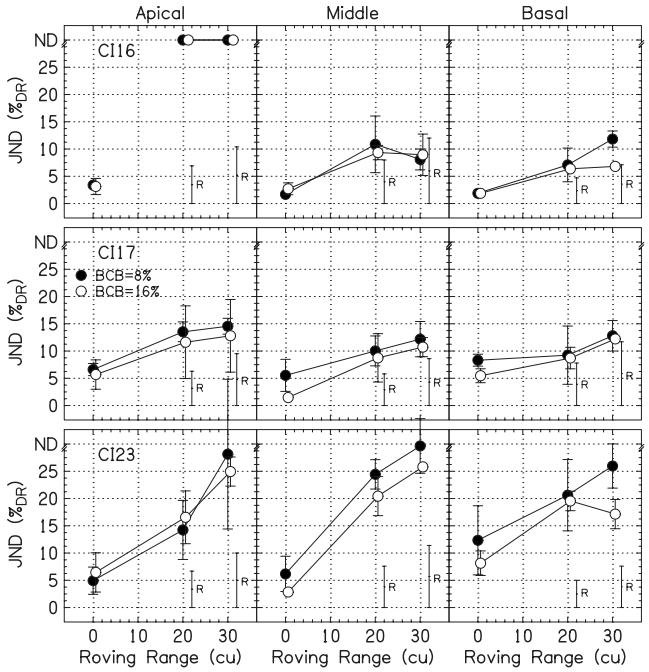


FIG. 8. Individual JNDs for peak height discrimination with overall level roving. The open symbols are for BCB=16%$_{DR}$ and the filled symbols are for BCB=8%$_{DR}$. The height of the roving bars, labeled "R," shows the level detection limit. JNDs that were unmeasurable are placed at the top of the panels at "ND." The error bars show ±1 standard deviation.

Our data on electrode discrimination for peaks can be compared to the data from electrode discrimination studies without a background. A study by Laneau and Wouters (2004) using Nucleus24 implants in monopolar mode found JNDs for electrode discrimination ranging from 0.34 to 0.61 electrodes for four individual subjects,[2] which translates to 0.26 and 0.46 mm, respectively. They observed no effect of bandwidth when changing the bandwidth from one to eight electrodes. Our mean JND for peaks was 2 mm, which is four to six times larger. However, several JND measurements were at the lower limit of measurement as given by the electrode spacings in the two studies, which confounds the comparison because the physical spacings of electrodes of the two implants are quite different. Nevertheless, the comparison for the majority of measurements not affected by floor effects indicates that adding a background makes electrode discrimination more difficult. This is in agreement with the results for NH listeners in Moore *et al.* where frequency discrimination of peaks and notches was found to be considerably worse than the discrimination of bands of noise or pure tones.

As for the previous experiments, notches were more difficult to discriminate than peaks and JNDs at the basal location were higher than JNDs at the apical and middle locations. Other studies have shown poorer electrode discrimination performance at basal locations without a background (e.g., Zwolan *et al.*, 1997; Pfingst *et al.*, 1999; Henry *et al.*, 2000).

V. EXPERIMENT 4: LARGE LEVEL ROVING

Once the parameter spaces for the three types of experiments were explored, selected conditions from the previous experiments were rerun with large amounts of overall level roving. The purpose of including large roving ranges was to clarify if CI listeners can perform profile analysis.

A. Methods

Experiments 1, 2, and 3 were performed again with overall level roving ranges of 0, 20, and 30 cu. Three of the listeners from the previous experiments participated: CI16, CI17, and CI23.

All three cochlear locations were used. Only peaks were used. For the repetition of experiment 1, peak detection, BW=2 and 3 were used. For the repetition of experiment 2, peak height discrimination, BW=2 and BCB=8 and 16%$_{DR}$ were used. For the repetition of experiment 3, electrode discrimination, BW=2 and 3 and BCB=12%$_{DR}$ were used. For the basal reference location, peaks were tested at apical target locations. For the middle and apical reference locations, peaks were tested at basal target locations.

B. Results

Figure 7 shows the data for peak detection from this experiment and experiment 1. Because there was no significant difference in the data between the experiments for the 0-cu roving range (p=0.48), they were combined. All of the data points are well above the level detection limit except one point, the BW=2, basal, 20-cu roving range for CI17, which was only slightly above the level detection limit.

Figure 8 shows the data for peak height discrimination with overall level roving. The JNDs for CI16 and CI17 are close to the level detection limit. For CI16, there are three data points below the level detection limit for the middle and

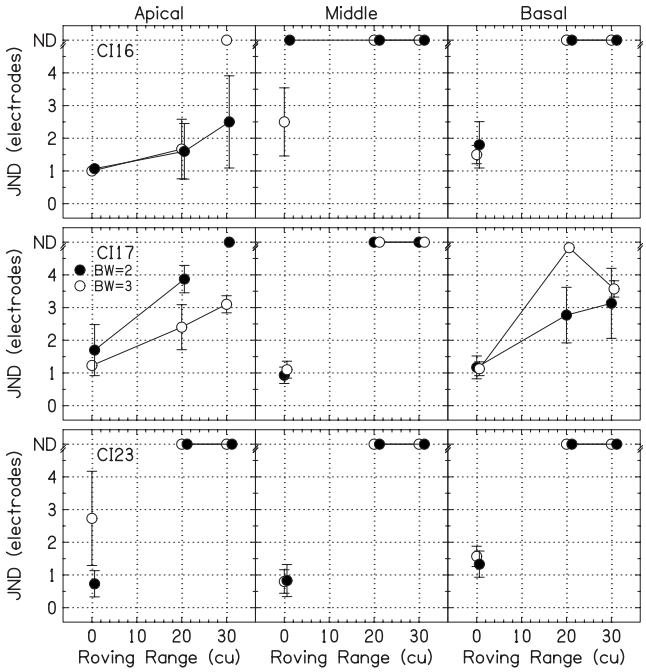


FIG. 9. Individual JNDs for electrode discrimination with overall level roving. The open symbols are for BW=3 and the filled symbols are for BW=2. JNDs that were unmeasurable are placed at the top of the panels at "ND." The error bars show ±1 standard deviation.

basal locations for the 30-cu roving range, but none statistically lower than the level detection limit. The JNDs for CI23 are well above the level detection limit.

Figure 9 shows the data for electrode discrimination with overall level roving. Most JNDs were unmeasurable. Only 28% of the JNDs could be measured with a nonzero roving range. The JNDs with roving are larger than those without roving. Note that, for these conditions, the use of within-channel cues cannot be excluded because the reference peak height of $12\%_{DR}$ was always larger than the level detection limit.

C. Discussion

Experiments 1, 2, and 3 were repeated with large amounts of level roving to check if listeners were performing intensity discrimination or profile analysis in these tasks. It appears that listeners were using only intensity cues because the JNDs were so large or because the JNDs could not be determined when roving was used. However, there are still some indications that profile analysis may be possible. In particular, the JNDs for peak height discrimination approached a level where they could not be achieved solely by monitoring the intensity within a channel. For NH listeners, Green and Kidd (1983) found that a "pedestal" minimizes masking effects while still allowing for profile analysis. In this task, when the BCB is larger, the across-channel interactions cause less masking of the target. Thus, these relatively small JNDs for peak discrimination compared to the JNDs for peak detection may be an indication of how severely the across-channel current spread affects the listeners' performance. Therefore, future profile-analysis experiments in CI listeners may want to focus on peak discrimination.

VI. GENERAL DISCUSSION

An important cue for NH listeners to detect changes in spectral profiles is the comparison of levels between bands (e.g., Green, 1988). One aim of this study was to determine whether CI listeners can do the same. CI listeners' sensitivity was measured to current increments and decrements on three or less electrodes presented with an equal-loudness background. The results showed many similarities in effects of cochlear location, bandwidth, and band current (level) with respect to the background between CI listeners and NH listeners; the analogous experiments with NH listeners were performed by Moore *et al.* However, in general, the data did not show that CI listeners could perform a profile analysis. The data showed a substantial increase in JNDs when roving was added, which is much different from most data for NH listeners, who show little effect of overall level roving. JNDs higher than 23.5% of the roving range can be achieved by using within-channel intensity cues only (Green, 1988, Appendix A), which is called the level detection limit. Most of our JNDs for conditions with roving were much higher than this limit. Also, intensity-discrimination JNDs for conditions without a background (band-only) were compared to JNDs for conditions with a background to see if there was an advantage due to the background, which has been shown for NH listeners (Green *et al.*, 1984; Green and Mason, 1985; Versfeld and Houtsma, 1991). In general, our data did not show a background advantage.

One question is why CI listeners cannot utilize across-channel comparisons to limit the detrimental effect of overall level roving while NH listeners can. It may be that CI listeners are inherently limited in some way. It is possible that channel interactions cause excessive blurring of the spectral shape, making it difficult to compare levels across different electrodes, which would explain many facets of our data. Drennan and Pfingst (2006) also performed an experiment using CI listeners, where the peak height on the center electrode was incremented while the number of electrodes and electrode density were varied. It was found that each electrode added to the background worsened the performance, even with large spacings between activated electrodes. (Note that this result may be confounded by the level being decreased to obtain a comfortable loudness each time more electrodes were added.) This is an interesting result, as NH listeners can show *decreasing* JNDs when energy outside the critical band is added for complex-tone stimuli, which indicates a background advantage. For NH listeners, the optimal log-spaced ratio of frequencies of the adjacent components is between 1.1 and 1.3 (Green, 1988). Denser spacings cause components to mask the target component and increase JNDs.

Kidd and Mason (1992) postulated that hearing-impaired (HI) listeners would be especially susceptible to level roving because of poorer frequency selectivity than NH listeners and loudness recruitment; these problems, in a way, are similar to those in CI listeners. Surprisingly, Lentz and Leek (2002, 2003) and Lentz (2006) showed no difference in performance between HI and NH listeners in profile-analysis tasks. The reason for this, as suggested by Lentz and Leek

(2003), may be that NH and HI listeners achieved similar performance by utilizing different parts of the spectrum to perform the task. The NH listeners may use frequencies near the spectral feature to be detected and the HI listeners may use frequencies remote from the spectral feature. It is possible that CI listeners, with even poorer frequency selectivity than HI listeners because of channel interactions, cannot benefit from using the edges of the spectrum and are thus hindered more than both NH and HI listeners in profile-analysis tasks.

Loudness is a potential cue in profile-analysis tasks and it is worth considering differences of loudness perception between acoustic and electric hearing. Loudness summation across electrodes in CI listeners is different from summation across frequencies for NH listeners. In models for acoustic hearing (Zwicker and Scharf, 1965; Moore and Glasberg, 1997), a signal is passed through a bank of overlapping bandpass filters to determine the cochlear excitation. This excitation is converted to specific loudness via a compressive power function. In a model for electrical hearing (McKay *et al.*, 2001, 2003), the excitation at each location in the cochlea is converted to specific loudness via an expansive power function because basilar membrane mechanics are bypassed in electrical stimulation. The overall loudness is calculated from the integration of specific loudness across cochlear location. As mentioned in the discussion of experiment 2, the loudness of components higher in level than the background (peak) will dominate the loudness perception, while components lower in level than the background (notch) will contribute negligibly to the loudness perception. Therefore by bypassing the compressive cochlear mechanics and altering the process by which loudness sums, the loudness could become the dominant perceptual feature for these stimuli, not the spectral shape.

It is possible that our lack of evidence for profile analysis in CI listeners is due to the specific properties of the stimuli used. There is evidence from the literature that temporal cues can contribute to profile analysis. Particularly for relatively narrowband spectra, pitch differences between the target and comparison sounds resulting from the evaluation of the temporal waveforms are important cues (e.g., Feth and Stover, 1987; Berg *et al.*, 1992). For very narrowband spectra, profile analysis may be mediated via detection of changes in the temporal envelope (Kidd and Mason, 1992; Green *et al.*, 1992). Encoding of spectral shape based on phase locking to temporal signal aspects seems to occur over the entire dynamic range, as has been shown with vowel stimuli (Young and Sachs, 1979). Hence, temporal cues seem to be resistant to overall level roving of the stimulus. The stimuli presented to the CI listeners in the present study contained no useful temporal cues, since all electrodes were stimulated at the same pulse rate. This may be a reason for the lower sensitivity of CI listeners in our profile-analysis task compared to NH listeners, particularly with level roving. The results of Moore *et al.* using noise stimuli, which provide less regularity than stimuli with discrete sinusoidal components, support this idea. At 1 kHz, where noise stimuli provide some temporal cues, they found higher JNDs compared to studies using tonal stimuli. At 8 kHz, where possibly neither type of stimulus provides useful temporal cues, they found no difference in JNDs. Moore *et al.* conjectured that this may also explain why level roving had a larger effect at 8 kHz than at 1 kHz. Interestingly, our data with level roving do not show more degradation of performance by level roving at the most basal place, consistent with the idea that the lack of temporal cues is the reason for the generally large JNDs with roving in electric hearing. Future CI studies could test the potential of temporal stimulus aspects in profile analysis with CIs by using stimuli that have temporal cues.

Several researchers have conducted profile-analysis experiments where the entire spectrum was altered, not just small portions of the spectrum (Bernstein and Green, 1987; Kidd and Mason, 1992; Lentz and Richards, 1997; Lentz and Leek, 2003). The advantage of such a technique is that it produces smaller JNDs, which reduces the roving range necessary to yield JNDs below the level detection limit. Such stimuli have not been used yet with CI listeners.

Since CI users can discriminate loudness-equated steady-state vowels (Laback *et al.*, 2004), this may indicate that they can perform some sort of profile analysis, ignoring possible within-channel cues. However, we did not test the electrodes that are probably most important for vowel discrimination. The tonotopic location of our lowest-frequency/most-apical target started at electrode 4, which normally receives information bandpass filtered between 700 and 900 Hz. It is possible that CI listeners can perform profile analysis when the peaks and notches are on the apical-most electrodes, electrodes 1–3. Continuing along this line of thought, there is no study known to the authors that tested spectral vowel discrimination with level roving. Based on the present study, it would not be surprising if vowel discrimination performance would decrease dramatically with roving.

Lastly and possibly the largest problem with our stimuli is that differences in loudness growth functions across electrodes could alter the background substantially from equal loudness when level roving was included. This is because we roved the overall level in cu instead of in loudness units. Kidd *et al.* (1986) showed that using stimuli that had a background with randomly varied amplitudes across frequency, frozen over 50 trials, increased JNDs by between 2.5 and 9.5 dB, depending on the range of the variation in the background. Larger increases in JNDs were seen if the background was varied from trial to trial. The variation of the background due to different growth of loudness curves at different electrodes may cause an effect similar to a random-amplitude background, which causes a large increase in JNDs. Future experiments could compensate for this limitation by using loudness growth curves for the level roving.

As mentioned before, Drennan and Pfingst (2006) performed a profile-analysis experiment with CI users, although they did not try to determine if the listeners were using an across-channel comparison as rigorously as we have. Their study did not show a background advantage. Henry and Turner (2003) and Henry *et al.* (2005) also tested CI listeners' sensitivity to spectral profiles with a spectral ripple reversal test. However, given that our results do not conclu-

TABLE II. The average JNDs over listeners from experiment 1 expressed as $\%_{DR}$, WF, and μA. Conditions for which no JND was measured are labeled "...".

Cochlear location		Apical			Middle			Basal		
BW (electrodes)		3	2	1	3	2	1	3	2	1
Peak	$\%_{DR}$	4.07 ± 2.56	4.80 ± 2.44	7.58 ± 4.26	4.12 ± 2.08	5.23 ± 2.82	10.56 ± 6.47	6.28 ± 3.58	6.91 ± 3.09	12.76 ± 6.32
No roving	WF (dB)	−7.62 ± 1.92	−6.75 ± 2.26	−5.12 ± 2.52	−7.91 ± 1.91	−7.12 ± 2.33	−4.58 ± 2.64	−6.89 ± 1.73	−5.71 ± 1.24	−3.35 ± 1.38
	μA	38.58 ± 22.85	42.17 ± 22.22	65.23 ± 43.93	34.55 ± 9.44	42.78 ± 18.93	78.65 ± 48.90	32.53 ± 14.71	37.10 ± 18.32	62.29 ± 32.53
Peak	$\%_{DR}$	10.74 ± 6.23	13.33 ± 5.61	16.93 ± 6.13	10.73 ± 4.72	12.27 ± 6.85	16.29 ± 9.43	11.27 ± 4.75	12.38 ± 5.06	17.65 ± 4.90
10-cu roving	WF (dB)	−4.03 ± 2.71	−2.59 ± 2.58	−1.48 ± 2.62	−4.33 ± 2.31	−3.83 ± 2.84	−2.59 ± 3.05	−4.36 ± 1.56	−3.27 ± 1.31	−1.51 ± 1.28
	μA	87.48 ± 63.51	99.23 ± 46.04	130.90 ± 69.12	78.73 ± 31.45	89.80 ± 51.30	117.95 ± 74.90	56.99 ± 26.96	68.16 ± 35.96	94.17 ± 55.28
Notch	$\%_{DR}$	6.83 ± 3.64	10.77 ± 7.89	9.88 ± 4.21	8.74 ± 5.06	7.22 ± 4.42	19.55 ± 11.81	20.99 ± 7.07	20.06 ± 12.69	26.44 ± 17.64
No roving	WF (dB)	−8.09 ± 3.60	−5.93 ± 3.19	−5.91 ± 3.37	−7.47 ± 4.60	−8.56 ± 4.53	−3.13 ± 5.01	−2.92 ± 2.16	−3.57 ± 3.42	−1.91 ± 2.65
	μA	35.80 ± 20.01	56.18 ± 43.92	62.36 ± 37.67	47.03 ± 29.53	37.54 ± 29.21	136.98 ± 94.52	62.56 ± 32.69	73.31 ± 65.79	112.57 ± 69.65
Notch	$\%_{DR}$	15.22 ± 7.86	…	22.09 ± 4.48	11.44 ± 0.81	…	…	18.25 ± 3.97	30.30 ± 3.23	…
10-cu roving	WF (dB)	−4.58 ± 3.10	…	−2.93 ± 1.06	−4.92 ± 0.38	…	…	−3.60 ± 1.10	−0.86 ± 0.55	…
	μA	95.48 ± 48.46	…	106.40 ± 23.13	139.95 ± 80.23	…	…	74.02 ± 17.78	131.84 ± 14.87	…
Band only	$\%_{DR}$	2.08 ± 1.27	2.95 ± 1.69	4.77 ± 3.47	3.52 ± 1.38	4.60 ± 1.93	8.79 ± 4.34	3.95 ± 2.24	4.76 ± 3.36	9.90 ± 5.71
	WF (dB)	−9.49 ± 1.37	−8.47 ± 1.32	−7.38 ± 2.04	−8.29 ± 1.40	−7.37 ± 1.66	−5.02 ± 2.30	−8.39 ± 1.66	−7.35 ± 2.02	−4.54 ± 2.41
	μA	23.09 ± 7.33	28.35 ± 11.15	39.95 ± 19.99	31.91 ± 8.65	38.65 ± 12.35	67.53 ± 33.42	27.78 ± 10.43	32.80 ± 16.43	61.21 ± 30.96

sively confirm that CI listeners are performing profile analysis, could the listeners in Henry *et al.* (2005) also have just monitored intensity within a single channel? The experiment used an 8-dB roving range, much smaller than the range necessary to eliminate the possibility of using intensity cues for a 30-dB peak-to-valley ratio. They admitted that the task could have been done without profile analysis. It seems that much more work is necessary to adequately address the question of profile analysis in electrical hearing.

VII. CONCLUSION

We assessed whether CI listeners can perform profile analysis. We know that CI listeners probably can perform some form of spectral-shape discrimination because they can discriminate vowels. However, our data showed no evidence of CI listeners performing across-channel comparisons in three basic types of profile-analysis experiments. We have suggested reasons for our results, both due to properties of CIs *per se* (spectral image blurring from channel interactions and different nature of loudness summation) and due to the specific stimuli used in this study (lack of temporal cues, stimuli where the entire spectrum is altered were not used, the lowest electrodes were not tested, and roving was not done in loudness units but in current units). Future experiments should consider these factors carefully.

ACKNOWLEDGMENTS

We would like to thank Mr. M. Mihocic for running experiments, our listeners, and the MED-EL Corp. for providing the equipment for direct electrical stimulation. We thank the associate editor, Brian Moore, and two anonymous reviewers for greatly improving this manuscript. This study was funded by the Austrian Science Fund (FWF Project No. P18401-B15).

APPENDIX

To increase the accessibility to the data, we provide all of the JNDs from experiments 1, 2, and 4 expressed as ten times the logarithm of the Weber fraction (WF) and as microamperes. The results of the experiments can be found in Tables II–IV. The WF was used by Pfingst *et al.* (1983) and Drennan and Pfingst (2006), and for CI listeners is

$$\mathrm{WF} = 10 \log[2\Delta I/I + (\Delta I/I)^2],$$

where I is the reference current and ΔI is the change in current. The WF has been shown to be an appropriate measure for electric stimulation since the variance across repeated measures at different base levels was shown to be constant (Nelson *et al.*, 1996). A majority of the results and all of the interpretations did not change using this measure,

TABLE III. The average JNDs over listeners from experiment 2 expressed as $\%_{DR}$, WF, and μA. Conditions for which no JND was measured are labeled "...".

Cochlear location		Apical	Basal
BW (electrodes)		2	2
Peak	$\%_{DR}$	3.19 ± 2.13	5.29 ± 2.46
+8$\%_{DR}$-BCB	WF (dB)	−9.43 ± 2.20	−7.88 ± 0.90
	μA	30.15 ± 20.74	36.54 ± 10.92
Peak	$\%_{DR}$	3.28 ± 1.96	5.46 ± 2.31
+16$\%_{DR}$-BCB	WF (dB)	−10.07 ± 1.91	−8.60 ± 1.05
	μA	29.79 ± 14.75	36.82 ± 8.54
Notch	$\%_{DR}$	11.64 ± 3.53	27.03 ± 1.61
−8$\%_{DR}$-BCB	WF (dB)	−6.08 ± 2.52	−1.98 ± 0.33
	μA	−67.35 ± 28.12	−103.59 ± 6.47
Notch	$\%_{DR}$	9.47 ± 1.69	…
−16$\%_{DR}$-BCB	WF (dB)	−7.83 ± 1.84	…
	μA	−59.72 ± 23.60	…

TABLE IV. The average JNDs over listeners from experiment 4 expressed as $\%_{DR}$, WF, and μA.

Cochlear location		Apical		Middle		Basal	
BW (electrodes)		3	2	3	2	3	2
Peak and Notch detection No roving	$\%_{DR}$	4.87 ± 1.56	6.43 ± 2.39	4.03 ± 1.97	4.82 ± 1.95	7.22 ± 4.10	7.77 ± 4.49
	WF (dB)	−6.45 ± 2.01	−5.53 ± 2.01	−7.67 ± 1.97	−6.85 ± 1.71	−6.12 ± 1.59	−5.75 ± 1.64
	μA	42.64 ± 14.57	54.60 ± 26.27	33.36 ± 6.72	39.01 ± 8.99	47.89 ± 16.21	50.64 ± 11.36
20-cu roving	$\%_{DR}$	17.79 ± 3.88	23.13 ± 3.61	14.31 ± 5.73	15.63 ± 4.31	14.27 ± 5.28	14.66 ± 8.28
	WF (dB)	−1.00 ± 2.77	0.28 ± 2.37	−2.83 ± 2.55	−2.13 ± 2.15	−2.99 ± 1.23	−2.87 ± 2.65
	μA	106.01 ± 17.85	135.82 ± 18.83	95.69 ± 25.84	104.41 ± 17.24	93.97 ± 36.78	105.98 ± 55.72
30-cu roving	$\%_{DR}$	23.14 ± 5.09	24.83 ± 1.93	18.43 ± 4.93	21.74 ± 7.31	18.55 ± 4.51	18.37 ± 4.84
	WF (dB)	0.24 ± 2.38	0.65 ± 2.09	−1.51 ± 1.71	−0.71 ± 2.29	−1.70 ± 0.83	−1.48 ± 1.36
	μA	136.21 ± 26.03	145.33 ± 10.59	124.93 ± 29.07	141.07 ± 34.36	125.04 ± 55.95	133.08 ± 62.04
Cochlear location		Apical		Middle		Basal	
BCB		+8$\%_{DR}$	+16$\%_{DR}$	+8$\%_{DR}$	+16$\%_{DR}$	+8$\%_{DR}$	+16$\%_{DR}$
Peak and notch Discrimination No roving	$\%_{DR}$	4.97 ± 1.99	5.08 ± 2.81	4.82 ± 2.91	3.75 ± 2.07	7.52 ± 5.60	5.16 ± 3.04
	WF (dB)	−7.78 ± 1.42	−8.82 ± 2.27	−8.45 ± 2.34	−9.92 ± 1.61	−7.27 ± 2.45	−8.73 ± 1.70
	μA	40.52 ± 10.84	39.99 ± 13.66	37.43 ± 11.79	33.20 ± 13.59	43.87 ± 12.89	36.55 ± 8.04
20-cu roving	$\%_{DR}$	13.88 ± 3.63	14.09 ± 5.87	16.28 ± 6.94	13.80 ± 6.36	12.31 ± 7.74	11.54 ± 6.24
	WF (dB)	−3.64 ± 1.49	−4.65 ± 2.66	−3.54 ± 2.28	−5.26 ± 2.60	−4.98 ± 2.69	−5.44 ± 2.09
	μA	81.47 ± 19.51	83.78 ± 30.42	106.39 ± 27.71	90.97 ± 25.70	77.93 ± 36.42	77.34 ± 23.82
30-cu roving	$\%_{DR}$	21.34 ± 11.48	18.88 ± 8.06	17.43 ± 9.98	15.85 ± 8.57	16.87 ± 7.31	12.06 ± 4.69
	WF (dB)	−1.98 ± 2.77	−3.37 ± 2.77	−3.46 ± 2.85	−4.79 ± 2.95	−3.23 ± 1.86	−5.10 ± 1.25
	μA	116.42 ± 45.99	107.55 ± 34.93	108.27 ± 30.84	100.97 ± 34.32	110.53 ± 48.65	81.41 ± 19.76

so it was omitted from the discussion. The data in microamperes are provided because they are a direct physical measure.

[1]The maximum DR from the group of two or three electrodes was also used to analyze the data, not just the average DR. It was found that the DRs were similar enough across different cochlear locations that a majority of the results and all of the interpretations did not change using this measure, so it was omitted.

[2]The fitting procedure used by Laneau and Wouters (2004) could lead to electrode discrimination JNDs less than 0.5 electrodes. The value of 0.34 corresponds to perfect pitch ranking and electrode discrimination in their task.

Berg, B. G., Nguyen, Q. T., and Green, D. M. (**1992**). "Discrimination of narrow-band spectra. I: Spectral weights and pitch cues," J. Acoust. Soc. Am. **92**, 1911–1918.

Berstein, L. R., and Green, D. M. (**1987**). "Detection of simple and complex changes of spectral shape," J. Acoust. Soc. Am. **82**, 1587–1592.

Donaldson, G. S., Kreft, H. A., and Litvak, L. (**2005**). "Cochlear location-pitch discrimination of single- versus dual-electrode stimuli by cochlear implant users," J. Acoust. Soc. Am. **118**, 623–626.

Drennan, W. R., and Pfingst, B. E. (**2005**). "Current-level discrimination using bipolar and monopolar electrode configurations in cochlear implants," Hear. Res. **202**, 170–179.

Drennan, W. R., and Pfingst, B. E. (**2006**). "Current-level discrimination in the context of interleaved, multi-channel stimulation in cochlear implants: Effects of number of stimulated electrodes, pulse rate, and electrode separation," J. Assoc. Res. Otolaryngol. **7**, 308–316.

Drennan, W. R., and Watson, C. S. (**2001**). "Sources of variation in profile analysis. II. Component spacing, dynamic changes, and roving level," J. Acoust. Soc. Am. **110**, 2498–2504.

Ellermeier, W. (**1996**). "Detectability of increments and decrements in spectral profiles," J. Acoust. Soc. Am. **99**, 3119–3125.

Feth, L. L., and Stover, L. J. (**1987**). "Demodulation processes in auditory perception," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 76–86.

Firszt, J. B., Koch, D. B., Downing, M., and Litvak, L. (**2007**). "Current steering creates additional pitch percepts in adult cochlear implant recipients," Otol. Neurotol. **28**, 629–636.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination*, Oxford Psychology Series No. 13 (University Press, New York).

Green, D. M., and Kidd, G. (**1983**). "Further studies of auditory profile analysis," J. Acoust. Soc. Am. **73**, 1260–1265.

Green, D. M., and Mason, C. R. (**1985**). "Auditory profile analysis: Frequency, phase, and Weber's law," J. Acoust. Soc. Am. **77**, 1155–1161.

Green, D. M., Berg, B. G., Dai, H., Eddins, D. A., Onsan, Z., and Nguyen, Q. (**1992**). "Spectral shape discrimination of narrow-band sounds," J. Acoust. Soc. Am. **92**, 2586–2597.

Green, D. M., Mason, C. R., and Kidd, G. (**1984**). "Profile analysis: Critical bands and duration," J. Acoust. Soc. Am. **75**, 1163–1167.

Heinz, M. G., and Formby, C. (**1999**). "Detection of time- and bandlimited increments and decrements in a random-level noise," J. Acoust. Soc. Am. **106**, 313–326.

Henry, B. A., McKay, C. M., McDermott, H. J., and Clark, G. M. (**2000**). "The relationship between speech perception and electrode discrimination in cochlear implants," J. Acoust. Soc. Am. **108**, 1269–1280.

Henry, B. A., and Turner, C. W. (**2003**). "The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 2861–2873.

Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.

Kidd, G., and Mason, C. R. (**1992**). "A new technique for measuring spectral shape discrimination," J. Acoust. Soc. Am. **91**, 2855–2864.

Kidd, G., Mason, C. R., and Green, D. M. (**1986**). "Auditory profile analysis of irregular sound spectra," J. Acoust. Soc. Am. **79**, 1045–1053.

Kidd, G., Mason, C. R., Brantley, M. A., and Owen, G. A. (**1989**). "Roving-level tone-in-noise detection," J. Acoust. Soc. Am. **86**, 1310–1317.

Koch, D. B., Downing, M., Osberger, M. J., and Litvak, L. (**2007**). "Using current steering to increase spectral resolution in CII and HiRes 90K users," Ear Hear. **28**, 38S–41S.

Kwon, B. J., and van den Honert, C. (**2006**). "Dual-electrode pitch discrimi-

nation with sequential interleaved stimulation by cochlear implant users," J. Acoust. Soc. Am. **20**, EL1–6.
Laback, B., Deutsch, W. A., and Baumgartner, W.- D. (**2004**). "Coding of vowellike signals in cochlear implant listeners," J. Acoust. Soc. Am. **116**, 1208–1223.
Laneau, J., and Wouters, J. J. (**2004**). "Multi-channel cochlear location pitch sensitivity in cochlear implant recipients," J. Assoc. Res. Otolaryngol. **5**, 285–294.
Lentz, J. J. (**2005**). "Profile analysis: the effects of rove on sparse spectra," J. Acoust. Soc. Am. **118**, 2794–2797.
Lentz, J. J. (**2006**). "Spectral-peak selection in spectral-shape discrimination by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 945–956.
Lentz, J. J., and Leek, M. R. (**2002**). "Decision strategies of hearing-impaired listeners in spectral shape discrimination," J. Acoust. Soc. Am. **111**, 1389–1398.
Lentz, J. J., and Leek, M. R. (**2003**). "Spectral shape discrimination by hearing-impaired and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 1604–1616.
Lentz, J. J., and Richards, V. M. (**1997**). "Sensitivity to changes in overall level and spectral shape: An evaluation of a channel model," J. Acoust. Soc. Am. **101**, 3625–3635.
McDermott, H. J., and McKay, C. M. (**1994**). "Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea," J. Acoust. Soc. Am. **96**, 155–162.
McKay, C. M., Henshall, K. R., Farrell, R. J., and McDermott, H. J. (**2003**). "A practical method of predicting the loudness of complex electrical stimuli," J. Acoust. Soc. Am. **113**, 2054–2063.
McKay, C. M., Remine, M. D., and McDermott, H. J. (**2001**). "Loudness summation for pulsatile electrical stimulation of the cochlea: Effects of rate, electrode separation, level, and mode of stimulation," J. Acoust. Soc. Am. **110**, 1514–1524.
Moore, B. C. J., and Glasberg, B. R. (**1997**). "A model of loudness perception applied to cochlear hearing loss," Aud. Neurosci. **3**, 289–311.
Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz,"J. Acoust. Soc. Am. **85**, 820–836.
Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (**1996**). "Intensity discrimination as a function of stimulus level with electric stimulation," J. Acoust. Soc. Am. **100**, 2393–2414.
Pfingst, B. E., Burnett, P. A., and Sutton, D. (**1983**). "Intensity discrimination with cochlear implants," J. Acoust. Soc. Am. **73**, 1283–1292.
Pfingst, B. E., Holloway, L. A., Zwolan, T. A., and Collins, L. M. (**1999**). "Effects of stimulus level on electrode-place discrimination in human subjects with cochlear implants," Hear. Res. **134**, 105–115.
Spiegel, M. F., and Green, D. M. (**1982**). "Signal and masker uncertainty with noise maskers of varying duration, bandwidth, and center frequency," J. Acoust. Soc. Am. **71**, 1204–1210.
Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). "Signal and masker uncertainty in intensity discrimination," J. Acoust. Soc. Am. **70**, 1015–1019.
Versfeld, N. J., and Houtsma, A. J. M. (**1991**). "Perception of spectral changes in multi-tone complexes," Q. J. Exp. Psychol. **43**, 459–479.
Young, E. D., and Sachs, M. B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibres," J. Acoust. Soc. Am. **66**, 1381–1403.
Zwicker, E., and Scharf, B. (**1965**). "A model of loudness summation," Psychol. Rev. **72**, 3–26.
Zwolan, T. A., Collins, L. M., and Wakefield, G. H. (**1997**). "Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **102**, 3673–3685.

Speech production variability in fricatives of children and adults: Results of functional data analysis

Laura L. Koenig[a)]
Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511 and Long Island University, One University Plaza, Brooklyn, New York 11201

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

Elizabeth Perlman[c)]
Rusk Institute, New York University Medical Center, 400 East 34th Street, RM 517, New York, New York 10016

(Received 3 February 2008; revised 29 July 2008; accepted 13 August 2008)

This study investigates token-to-token variability in fricative production of 5 year olds, 10 year olds, and adults. Previous studies have reported higher intrasubject variability in children than adults, in speech as well as nonspeech tasks, but authors have disagreed on the causes and implications of this finding. The current work assessed the characteristics of age-related variability across articulators (larynx and tongue) as well as in temporal versus spatial domains. Oral airflow signals, which reflect changes in both laryngeal and supralaryngeal apertures, were obtained for multiple productions of /h s z/. The data were processed using functional data analysis, which provides a means of obtaining relatively independent indices of amplitude and temporal (phasing) variability. Consistent with past work, both temporal and amplitude variabilities were higher in children than adults, but the temporal indices were generally less adultlike than the amplitude indices for both groups of children. Quantitative and qualitative analyses showed considerable speaker- and consonant-specific patterns of variability. The data indicate that variability in /s/ may represent laryngeal as well as supralaryngeal control and further that a simple random noise factor, higher in children than in adults, is insufficient to explain developmental differences in speech production variability. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2981639]

PACS number(s): 43.70.Ep, 43.70.Aj [BHS] Pages: 3158–3170

I. INTRODUCTION

Many past studies have demonstrated decreasing token-to-token variability in speech and other motor acts with increasing age, but interpretations of this finding have varied widely (see Secs. I B and I C). Some authors have argued that immature physiology yields more random sensorimotor noise in children than in adults, whereas others have emphasized patterned variation reflecting individual learning and exploration in the context of ongoing anatomical or physiological change. This study explored the characteristics of speech production variability in children using two strategies. First, variability was compared across two articulator systems, the larynx and the tongue, by analyzing oral airflow data in the fricatives /h s z/. This methodology yields information on the timing and magnitude of laryngeal and lingual actions in a noninvasive fashion. Second, variability in repeated productions of these consonants was quantified using functional data analysis (FDA). FDA is a useful tool for this purpose because it allows calculation of relatively independent measures of temporal and amplitude variabilities. These results may provide a more detailed understanding of how speech motor control develops across articulatory systems and in temporal as well as spatial domains. A fuller description of developmental variability may, in turn, speak to the underlying sources of this phenomenon.

The following sections review (a) articulatory, aerodynamic, and developmental characteristics of fricatives; (b) variability in children's speech and other motor behaviors, considering (c) issues of methodology and interpretation; and (d) the characteristics of FDA and its appropriateness for the current work.

A. Fricative production and development

The focus on /h s z/ in this work was motivated by two considerations. Most fundamentally, these three fricatives, in combination, provide insight into laryngeal, oral, and combined laryngeal-oral activities: /h/ requires vocal fold abduction but no particular supralaryngeal movements; /z/ requires a tongue-tip constriction, while the vocal folds remain adducted for voicing; and /s/ requires both laryngeal abduction for devoicing and a tongue-tip constriction. A useful feature of oral airflow data is that they permit noninvasive assessment of both articulatory actions: Airflow rises with vocal fold abduction and decreases with tongue-tip constrictions.

[a)]Author to whom correspondence should be addressed; electronic mail: koenig@haskins.yale.edu
[b)]Electronic mail: lucero@mat.unb.br
[c)]Electronic mail: elizabeth.perlman@nyumc.org

 © 2008 Acoustical Society of America

In the case of /s/, the signals typically show an increase related to abduction and a superimposed valley reflecting lingual actions (Klatt *et al.*, 1968).

Further, from a developmental perspective, the fricatives /s z/ are of interest because they are acquired rather late, on average, for many children (Sander, 1972). Kent (1992) proposed that lingual fricatives are late-appearing sounds because they require fine control of tongue position and force. Yet fricatives also have specific aerodynamic requirements; in particular, they require sufficient airflow to generate turbulent noise (Howe and McGowan, 2005, Scully *et al.*, 1992, and Shadle, 1990). Indeed, direct laryngeal data show that adult speakers may use a slightly more open glottal configuration in /z/ compared to surrounding vowels, presumably to increase airflow a bit without actively inhibiting voicing (Lisker *et al.*, 1969; Sawashima, 1970). If lingual fricatives are difficult for children because of their supraglottal requirements, one would expect to see parallel patterns for /s/ and /z/ within individuals. On the other hand, the combination of a precise supraglottal constriction and abduction for devoicing and high airflow in /s/ may yield an added level of complexity compared to /z/.

There has been surprisingly little discussion of laryngeal function in /s/ acquisition. This is in marked contrast to the long history of studying children's laryngeal control for stop consonant voicing, especially via measures of voice onset time (VOT) [e.g., see Kewley-Port and Preston (1974) and Macken and Barton (1980)]. A consistent finding in the VOT literature has been that adultlike levels of variability are reached later for voiceless aspirated stops than for their unaspirated counterparts. The traditional explanation for this has been that voiceless aspirated stops are challenging for children because they require precise temporal coordination of vocal fold abduction and oral release (Kewley-Port and Preston, 1974). The data obtained by Koenig (2000) suggested an alternative explanation. That study compared the duration of devoiced regions in /h/ with VOTs in English aspirated /p t/ produced by 5 year olds and adults, and also assessed the degree of abduction in /h/ using oral airflow signals. The results showed significant correlations between devoiced regions in /h/ and /p t/ VOTs, as well as great variability in the extent and duration of abduction in /h/ for many children. If children are highly variable in their abduction patterns for /h/—a consonant that requires *only* a laryngeal gesture—then it is likely that their laryngeal control will also be variable for consonants that require abduction along with supraglottal articulations. The current work extends that logic to the study of voiceless fricatives. Specifically, comparing variability patterns in /s/ to those of /h/ and /z/ can help elucidate how /s/ reflects laryngeal as well as oral actions and, more generally, shed light on the degree to which developmental variability is articulator specific, both within and across individuals.

B. Variability in development

Children, as a group, demonstrate greater token-to-token variability than adults when repeating the same speech sequence for acoustic, aerodynamic,[1] electropalatographic, and kinematic measures [e.g., see Cheng *et al.* (2007), Chermak and Schneiderman (1986), Eguchi and Hirsh (1969), Kent (1976), Kent and Forner (1980), Munson (2004), Ohde (1985), Sharkey and Folkins (1985), Smith and Goffman (1998), Smith and McLean-Muse (1986),Tingley and Allen (1975), and Zajac and Hackett (2002)]. Child-adult differences can persist well into the school-age years (Eguchi and Hirsh, 1969; Kent and Forner, 1980; Lee *et al.* 1999; Munson, 2004; Ohde, 1985; Sharkey and Folkins, 1985; Smith and McLean-Muse, 1986; Zajac and Hackett, 2002) and even into adolescence (Walsh and Smith, 2002). Similar age-related effects have been reported in studies of nonspeech motor tasks such as reaching, tapping, and pinching (Berthier, 1996; Deutsch and Newell, 2001, 2004; Tingley and Allen, 1975; Yan *et al.*, 2000), suggesting that the phenomenon largely reflects motoric (rather than linguistic) development. This paper will mostly treat speech production variability from a motor control perspective, but results from work such as this may have implications for theories of developmental phonology, and some results from more linguistically oriented work will be addressed briefly below.

Several factors must be considered when evaluating the nature of developmental variability. One issue is that children's motor acts are not only variable on average relative to adults' but also tend to display lower velocities and longer segment or movement durations (e.g., see Chermak and Schneiderman (1986), Eguchi and Hirsh (1969), Kent and Forner (1980), Smith and Goffman (1998), Smith (1978), and Yan *et al.* (2000)]. Since variability, all else being equal, increases with the absolute magnitude of a measure (Ohala, 1975), developmental studies often employ relative measures such as the coefficient of variation, where the standard deviation (SD) is divided by the mean. In a series of studies, Smith and co-workers demonstrated that longer durations account for part but not for all of the higher variabilities seen in temporal measures of children's speech [Smith, 1992; Smith, 1994; Smith *et al.*, 1983; Smith *et al.*, 1996; see also Chermak and Schneiderman (1986)].

Further, age-related differences may vary depending on the articulators and signal types assessed. For example, Green *et al.* (2002) reported that 1- and 2-year-old children have more adultlike variability for the jaw than for the lips. Nittrouer (1993) similarly proposed that young children have more mature control over the jaw than the tongue. Stathopoulos (1995) emphasized that high variability observed at one level of analysis does not necessarily imply high variability on others, even for measures typically thought to be closely related, such as subglottal pressure and speaking sound pressure level. Goffman *et al.* (2002) presented another example of differences across signal types. In a longitudinal case study on speech production after cochlear implantation, these authors found that durational variability in the acoustic signal declined steadily following implantation, whereas kinematic variability first increased before decreasing to more age-appropriate levels. Although these changes over time presumably resulted from the child's changing auditory feedback rather than simple speech motor development, they do indicate that kinematic and acoustic recordings of the same speech events may show different trends.

Smith and Goffman (1998) considered whether temporal and spatial aspects of speech mature at the same rate. Their study used the spatiotemporal index (STI) to quantify lower-lip displacement variability in 4 year olds, 7 year olds, and adults producing repetitions of "buy Bobby a puppy." As expected, STI results were significantly higher in children than in adults. Since the STI combines the temporal and amplitude aspects of variability into a single number, two subsequent analyses were performed on the time-normalized data to investigate possible differences between these two dimensions. An assessment of the relative timing of velocity peaks in the lowering movements for /ɑ/, /iə/, and /ʌ/ in "Bobby a pup-" revealed a tendency, significant for the 7 year olds but not for the 4 year olds, for children to produce their first and second velocity peaks relatively later in the utterance than adults. In contrast, success rates for a cross-correlation pattern-matching classification on the lower-lip displacement in the syllables "Bob" and "pup" showed no significant age effects. Smith and Goffman (1998) concluded that relative timing in an utterance matures later than spatial displacement in individual syllables. The authors characterized this finding as unexpected since the long time course of physical growth might lead one to expect spatial features to show more protracted development than temporal features. It is interesting to note in this context, however, that a longer acquisitional time frame for temporal aspects of speech production is reminiscent of a theme that has recurred in the developmental phonology literature. Specifically, some child language researchers (Ferguson and Farwell, 1975; Vihman, 1993; Waterson, 1971) have observed that young children's word productions may contain many articulatory gestures appropriate for the adult model but lack consistent temporal phasing between them.

The methods of the present study were designed to address many of the issues identified in the preceding paragraphs. Specifically, variability measures in FDA were derived relative to the mean and SD for each speaker and consonant; variability was compared across different articulators; and measures were obtained for both temporal and spatial aspects of the data.

C. Interpreting variability

The trends of decreasing motoric variability, higher movement velocities, and shorter durations with age have often been attributed to neurological changes including cortical and cerebellar development, stabilization of electroencephalography (EEG) patterns, and myelination of the motor tracts [e.g., see Kent (1976), Lecours (1975), Tingley and Allen (1975), and Yan *et al.* (2000)]. From this perspective, variability in children may result from increased sensorimotor noise in a less stable system. Contemporary models of motor control and sensorimotor integration often include explicit noise terms [e.g., see Körding and Wolpert (2004), Todorov and Jordan (2002), and Vetter and Wolpert (2000)] and Berthier (1996) successfully modeled infant reaching data using a high level of random noise.

Other work suggests that random neurological noise may not differ as a function of age, however. First, extensive cross-subject differences have been observed in developmental studies, with some children having rather adultlike values for some variability measures (Smith, 1995; Smith *et al.*, 1996; Smith and McLean-Muse, 1986; Stathopoulos, 1995). Such individual variation would seem to argue against a simple additive noise factor that differs between children and adults (Stathopoulos, 1995). Further, Deutsch and Newell (2001, 2003, 2004) found that frequency analyses of finger pinching did not show age differences in random noise; rather, children's movements had reduced dimensionality and more energy in lower frequencies. These authors proposed that children differ from adults in being less skilled at meeting task demands and integrating sensory feedback during ongoing movement control.

The motor learning literature indicates that practicing a task leads to faster, more accurate, and less variable performance (Hatze, 1986; Liu *et al.*, 2006; Schulz *et al.*, 2000; Sosnik *et al.*, 2004). Thus, similar developmental patterns for speech durations and variability (Eguchi and Hirsh, 1969; Kent, 1976) may simply result from children being less practiced speakers. Deutsch and Newell (2004) observed that, with practice, some children approached adultlike values of force variability in a nonspeech (pinching) task and, moreover, that practice effects accounted for more variance than chronological age. Age effects can also be reduced by manipulating task demands, including accuracy requirements, force levels, and presence of feedback (Deutsch and Newell, 2001, 2003, 2004; Yan *et al.*, 2000).

Finally, both speech and nonspeech studies indicate that production variability may increase during periods of learning or skill acquisition when the motor system is in transition to a new mode of organization (Berthier, 1996; Goffman *et al.*, 2002; Liu *et al.*, 2006; Thelen *et al.*, 1996; von Hofsten, 1989). In this view, specific patterns of variability arise as individuals seek successful motoric strategies for a given goal (Jensen *et al.*, 1995). Goffman *et al.* (2002) [see also Smith and Goffman (1998)] noted that greater plasticity in the developing nervous system may facilitate task-space exploration and be a sign of a flexible and adaptive organization. This implies that the underlying causes of variability may have functional utility during development.

Although variability may ultimately derive from many sources—and this can lead to some indeterminacy [see Hatze (1986)]—sophisticated experimental paradigms may yield greater insight into the likely sources of developmental variability for particular tasks, as well as the nature and time course of age-related changes. As outlined in the next section, FDA allows token-to-token variability to be partitioned into spatial and temporal domains. This, combined with a comparison across articulators, may speak to the general debate about the extent to which developmental variability reflects a general random noise factor versus individualistic task-specific patterns.

D. FDA as a technique for exploring variability

FDA was developed by Ramsay and Silverman (1997) as a means of analyzing time-varying signals. A distinctive feature of FDA is nonlinear time warping to bring multiple

signals into closer alignment with an average signal. These methods may be used in service of many questions. For example, previous researchers have used FDA to explore the underlying dimensionality of speech movements (Ramsay *et al.*, 1996), how speech timing changes near phrasal boundaries (Lee *et al.*, 2006), the most appropriate method for calculating the harmonics-to-noise ratio of voice signals (Lucero, 1999), and the degree to which kinematic variability in consonants depends on the articulators involved in forming the constriction (Lucero and Löfqvist, 2005). The present work capitalizes on the fact that FDA allows variability to be decomposed into two dimensions: The magnitude of the time warping functions used to align the data provides an index of temporal (phasing) variability, whereas the remaining variance in the aligned waveforms represents amplitude variability once most phasing variability has been factored out. One question of this work was to determine whether these two forms of variability show parallel developmental trends.

E. Current study

In summary, the purpose of this study was to investigate speech production variability in children using two strategies. First, variability was explored across oral, laryngeal, and combined oral-laryngeal actions. It may be that stable control is acquired earlier over one of these articulatory systems than the other. Alternatively, speakers may differ in which system shows more stable behavior at a particular time. Comparing the variability patterns for the three consonants was intended to elucidate the ways in which children, both individually and as a group, learn to control lingual and laryngeal articulations alone and in combination. Variability in the three consonants was assessed quantitatively using FDA methods and qualitatively, over time, in a vowel-consonant-vowel (VCV) sequence. Second, decomposing variability into distinct phasing and amplitude components permitted exploration of whether these two dimensions have similar developmental profiles. *A priori*, it seemed sensible to predict that both phasing and amplitude variabilities would be higher in children than in adults, but whether the two measures have equivalent developmental time courses was an open question. Differences in variability patterns for the three consonants or for temporal and amplitude variabilities would suggest that a simple random noise factor is insufficient to explain variability differences between children and adults. Thus, the specific questions of this study were as follows: (a) How does variability compare across /h s z/ within and across age groups? (b) Do amplitude and phasing variabilities develop in parallel for this set of consonants?

II. METHODS

A. Speakers

Ten children in each of two age groups and ten women were recorded. The younger children ranged in age from 4.5 to 6.0 years (μ=5.07, henceforth called 5 year olds). The older children were 8.9–10.8 years of age (μ=9.86, henceforth called 10 year olds). The child age groups were chosen so as to obtain data from children (a) as young as possible, given the demands of the experimental task, and (b) as old as possible without introducing gender differences as a function of puberty or minimizing the likelihood of observing age effects. Some measures of variability in children may approach adultlike levels at around 12 years of age (Eguchi and Hirsh, 1969; Lee *et al.*, 1999). Each of the child groups had five boys and five girls.

All participants were native speakers of American English from the New York City metropolitan region. The children had normal developmental histories as ascertained by parental report, including no history of intervention for speech or language problems. Parents of the 10 year olds also verified that their children had not yet begun to show pubertal changes. A short narrative or conversational sample was obtained from each child speaker (targeting 50 or more utterances; all included at least 35) and was analyzed to confirm age-appropriate language skills, as measured by mean length of utterance (MLU) and syntactic complexity (Hughes *et al.*, 1997; Miller and Chapman, 1981). Finally, all children passed a hearing screening to establish binaural thresholds at 25 dB or less over 250–8000 Hz.

The adult female group consisted of mothers of the child participants. All had normal speech, language, and hearing histories by self-report and speech characteristics within normal limits as judged informally by the first author. Mothers were used as the comparison group for two reasons. First, some of the younger children were more comfortable doing the experiment after watching their mothers perform the task. Second, this minimized dialectal variation, an important consideration with relatively small group sizes. The adult group was further restricted to women because they have laryngeal and vocal-tract proportions more similar to children's than do men. Although restricting the adult group to women may limit the generalizability of the adult data somewhat, no evidence exists to indicate that men and women differ in their variability characteristics for repeated productions of speech sequences, so this limitation should not affect general conclusions regarding age effects.

The experimental protocol was approved by an Institutional Review Board and followed standard procedures for the protection of human research participants. Before recording, parents provided informed consent for their own and/or their child's participation, and all children provided assent. All participants were naive as to the purposes of the experiment.

B. Speech materials

Speakers were recorded producing approximately 25–30 repetitions of the utterances "Poppa Hopper," "Poppa Sapper," and "Poppa Zapper." The targets /h s z/ thus had a common preceding context (bilabial closure followed by schwa) and a low vowel following (/a/ or /æ/). The utterances were elicited by means of colored drawings of a bunny carrying out activities related to his name (viz., hopping, draining sap from a tree, and getting zapped by a lightning bolt). The target consonant was also prominently displayed on the bunny's shirt, and his name was printed at the top or bottom of the picture. A research assistant introduced the

pictures to speakers before recording, pointing out the salient features of the picture. For the younger (preliterate) child group, picture naming was elicited immediately following the description (e.g., "This guy is getting SAP from a tree, so he's Poppa SAPPER. Can you say that?"). Speakers were told that when they saw a picture, they should say the bunny's name and then repeat it five times.

During recording, the research assistant presented pictures for naming and counted off on her fingers as the speaker repeated the utterance four to five times. Since the utterance list was very short, the target consonant was given on the bunny's shirt, and the bunny's name was written on the picture, all adults and most child speakers were usually successful in naming the picture accurately on the first presentation. For some of the 5 year olds, additional prompts were given if the child's first production was not perceptually adequate (e.g., "He's hopping. What's his name?"). Pictures were presented five times per recording session, in randomized order. For analysis, tokens were excluded if there was a pause before the target consonant or if the target consonant was not perceptually accurate in place, manner, and voicing. No speaker was included in the final data set if s/he did not produce at least ten analyzable tokens of each consonant. The average number of tokens analyzed for each consonant and speaker was 27 (SD=5).[2]

Most often the speakers produced all repetitions on a single breath, but a few of the 5 year olds occasionally inspired between repeated productions. Speech rate was not explicitly controlled, but any speaker who at first used a slow careful speaking style was encouraged to speak more naturally and casually so that there were no perceptually clear pauses between words within a target utterance.

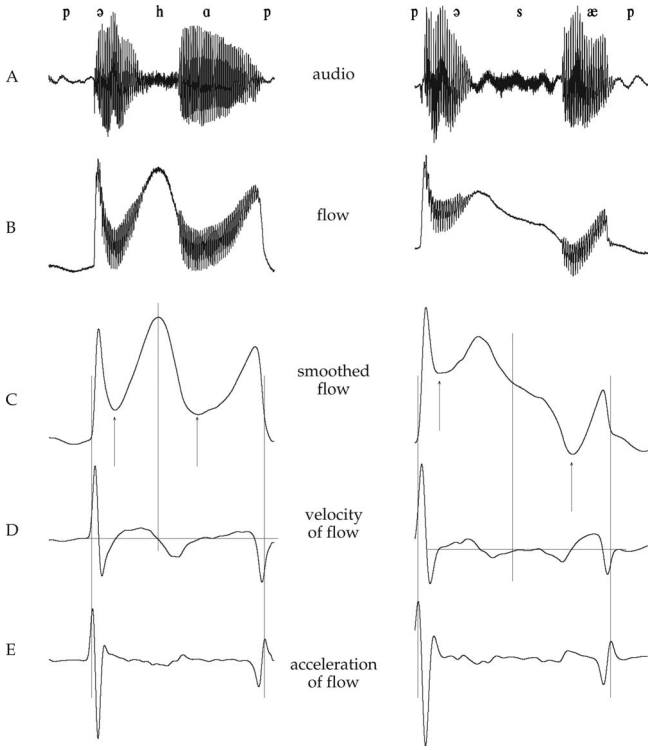


FIG. 1. Sample tokens of /pəhap/ and /pəsæp/ from a 5-year-old boy. The time axis is 500 ms for both sequences. (A). Audio signal. (B) Original airflow signal. (C) Smoothed airflow signal. The arrows show the region over which the final amplitude index was calculated (viz., between the flow minima of the two vowels). (D) Velocity of smoothed airflow. Zero-crossings in this signal were used to identify flow peaks in /h/ and minima in /s z/. (E) Acceleration of smoothed airflow. The peaks in this signal reflecting release and closure of the preceding and following /p/s marked the boundaries of the FDA warping window.

C. Equipment and recording

Oral airflow signals were collected using an undivided Glottal Enterprises pneumotachograph appropriately sized for the speaker. The airflow data were digitized at 10 kHz onto a PowerLab recording system connected to a Macintosh laptop computer. An audio signal was simultaneously recorded to the PowerLab at a sampling rate of 20 kHz using an AKG 420 microphone hung around the speaker's neck. The microphone signal was used to determine that the target consonant was accurate at the level of broad transcription and that the speaker did not pause before the target consonant. All analysis for this work was performed on the airflow data. Flow signals were calibrated before or after each recording session using a rotameter.

D. Preliminary processing

Initial data processing used smoothing and differential routines in the Chart™ software that accompanies the PowerLab. Since the articulatory activities of interest, namely, glottal abduction and supraglottal constrictions, are relatively low in frequency, the airflow data were smoothed twice using a triangular window of 201 points (approximately 20 ms, or a 50 Hz low-pass cutoff). This removed the rapid oscillations associated with vocal fold vibration. First and second derivative (velocity and acceleration) signals were obtained from the smoothed flow signals and smoothed iteratively with 201-point windows (same specifications as above) so that peaks and zero-crossings could be obtained easily. Typically two passes of filtering were sufficient for this purpose, but for a few speakers whose signals were noisier, a third round of filtering was implemented to remove remaining high-frequency oscillations that complicated the semiautomatic event selection. Individual tokens of /h s z/ were then labeled as follows (see Fig. 1): The peak of the smoothed flow signal in /h/, representing the maximum abduction, was identified from zero-crossings in the velocity signal. Flow minima during the oral constrictions for /s z/ were likewise identified using the velocity zero-crossing. Finally, the second-derivative peaks representing the flanking /p/ release and closure were identified. All peak and zero-crossing labeling was performed with reference to the original flow signals to ensure that appropriate locations were chosen. To prepare the data for FDA processing, individual tokens were extracted by selecting a window centered around the /h s z/ peak or valley, long enough so that it included both /p/ acceleration peaks. These signals (smoothed flow data) were saved as text files and imported into MATLAB. In MATLAB, the data were trimmed so that each token began and ended with consistent aerodynamic events, namely, the second-derivative peaks reflecting /p/ release and closure. This defined the VCV sequence used in subsequent processing.

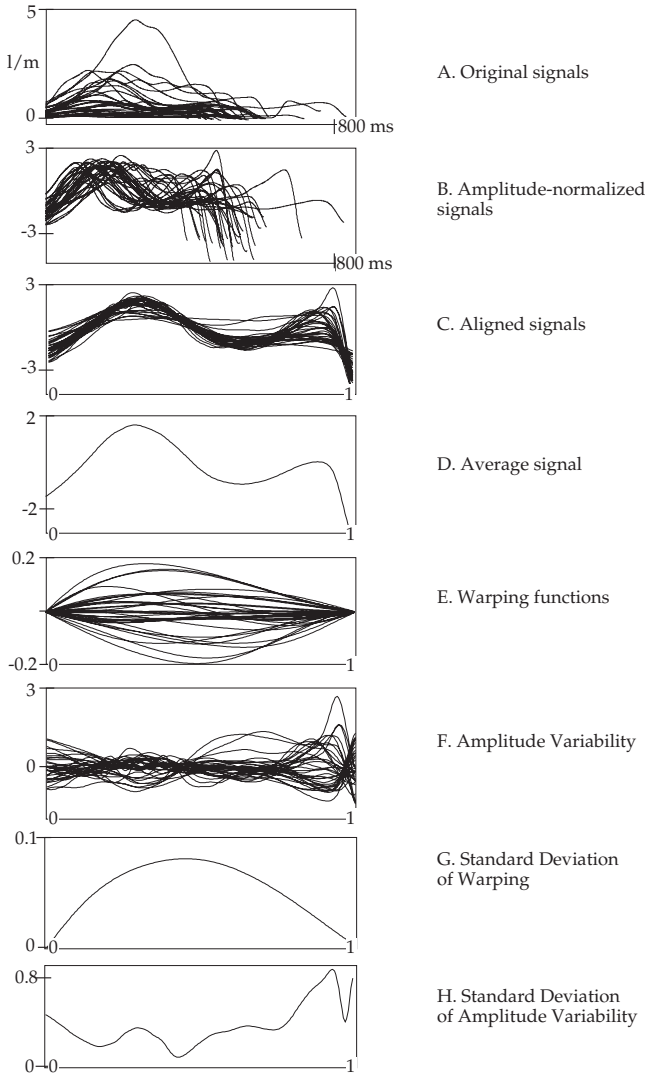


FIG. 2. An illustration of the processing steps for all tokens of /h/ in speaker 5m1 (a 5-year-old boy). Panels top to bottom: (A) All original productions of /h/ not normalized for time or amplitude. (B) Amplitude-normalized tokens. After amplitude normalization, the *y*-axis is unitless. (C) Time-normalized (i.e., aligned) tokens. After time normalization, the *x*-axis is unitless. (D) Average of all tokens after time and amplitude normalization. (E) Warping functions for all tokens. (F) Amplitude variability (SDs) for all original records relative to average. (G) SD of warping functions over normalized time. (H) SD of amplitude variability functions over normalized time.

E. FDA processing

FDA methods may vary somewhat depending on the goals of the work, but the general purposes are to reduce the data (so as to facilitate processing) and, most critically, to align the signals in time using nonlinear time warping. The magnitude of the time warping required to bring the set of signals into alignment (as determined by a cost-minimization function) represents phasing variability, and the variation around the average obtained from the aligned data provides a measure of amplitude variability, which is relatively independent of temporal variability. The specific processing steps used in the current work, for all tokens of a consonant within a speaker, were as follows (see Fig. 2):

(a) The original tokens [panel (A) in Fig. 2] were vertically aligned by subtracting the mean of each token and amplitude normalized by dividing each token by its SD [Fig. 2(B)]. This normalization corrects for greater variability resulting from increased magnitudes alone (see Ohala, 1975).

(b) The amplitude normalized tokens were processed using the FDA package of MATLAB functions given by Ramsay and Silverman at ftp://ego.psych.mcgill.ca/pub/ramsay/FDAfuns [see also Ramsay and Silverman (1997, 2002)].

First, the tokens were put into functional form by expanding them into a basis of 42 fourth-order B-splines. Using an order of 4 ensures that the signals will produce smooth acceleration functions. At the same time, their duration was set to a normalized time span of [0, 1]. This number of basis functions was sufficient to capture details of the tokens' waveshapes, while applying a light smoothing, as assessed by a visual inspection of the results.[3]

The tokens were then aligned in time by nonlinear time warping. In this process, nonlinear transformations of time $h(t)$ (warping functions) are computed for each token so as to minimize the following measure of shape similarity:

$$\mathrm{MINEIG}_{\lambda}(h) = \mathrm{MINEIG}(h) + \lambda \int w^2(t)dt, \tag{1}$$

where w is the relative curvature of the warping function, λ is a roughness penalty coefficient imposed on the warping function, and MINEIG(h) is the logarithm of the smallest eigenvalue of matrix

$$\begin{bmatrix} \int [x_0(t)]^2 dt & \int x_0(t)x[h(t)]dt \\ \int x_0(t)x[h(t)]dt & \int x[h(t)]^2 \end{bmatrix}, \tag{2}$$

where $x(t)$ is the token being normalized and $x_0(t)$ is the target for the time warping. For these analyses, λ was set to 10^{-5}, ensuring smooth warping functions of time. Further, the warping functions were expressed in a basis of ten fourth-order B-splines, following previous work (Lucero *et al.*, 1997). The computation of the optimal warping functions was performed in two steps. A first approximation was obtained using the mean of the tokens as the target [$x_0(t)$ in Eq. (2)]. A new target was then obtained as the mean of the aligned tokens (computed in the first step), and the warping functions were refined in a new run of the algorithm. Panel (C) in Fig. 2 shows the data after applying the time warping, and panel (D) shows the average calculated from the aligned data.

(c) To assess the degree of temporal variability of the signals, the SD of the time warping functions over time was obtained [see panels (E) and (G)].

(d) To assess the amount of amplitude variability in the aligned signals, the waveform noise (deviation from the average signal) was obtained for each token [panel (F)], and the SD of this set of functions was obtained [panel (H)].

Early analyses indicated that speakers (women as well as children) tended to have high variability at the beginning and/or end of their records [note the late peak in the amplitude SD function in Fig. 2 (H)], resulting from airflow variation related to the preceding and following /p/. Since the goal of this analysis was to characterize the VCV sequence and not the aspects of /p/ *per se*, the final amplitude index was calculated over the region defined by flow minima in the vowels flanking the target consonant. This region is indicated by arrows in Fig. 1(C).

The final analyses were based on the following data for each consonant and speaker: (a) Average smoothed airflow signals for /h s z/. (b) Two SD functions over time (amplitude and time warping). (c) The mean of the time warping SD [i.e., the mean of the signal in Fig. 2(G)] obtained over the entire VCV sequence (henceforth called the time warping index). (d) The mean of the SD of the waveform noise [Fig. 2(H)] calculated over the region indicated in Fig. 1 (henceforth called the amplitude index).

The time warping index provides a measure of temporal variability in the airflow peaks and valleys across tokens. The amplitude index represents the degree to which the magnitude of oral flow varied across tokens after factoring out temporal variability. Flow rates, in turn, reflect laryngeal and oral events: Laryngeal abduction will increase airflow rates, all else being equal, whereas supraglottal constrictions will reduce airflow. Thus, the two indices measure different types of production variability: gestural phasing versus magnitude of abduction or tongue-tip constriction. These summary measures permit statistical testing of age and consonant effects on variability. The amplitude and time warping functions over time were also assessed qualitatively to gain insight into individual patterns of variability.

F. Statistics

The amplitude and time warping indices for the 3 consonants and 30 speakers were entered into two-way analyses of variance (ANOVAs) using STATVIEW, with independent variables of age group and consonant. Since two separate ANOVAs were being run, the α-level was set to 0.025. In cases of significant main effects or interactions, *post hoc* Fischer's PLSD tests were run. Correlations (Pearson's r) were also carried out to evaluate the relationship between amplitude and time warping indices within speakers.

III. RESULTS

A. Age and consonant effects on amplitude and time warping indices

As indicated above, the amplitude and time warping indices are useful for performing statistical tests of how variability differs as a function of age, across consonants, and in temporal and amplitude domains. Figure 3 presents the amplitude (top) and time warping (bottom) values for all speakers. The average amplitude and warping indices are given for the three age groups in Fig. 4. Tables I and II summarize the results of the ANOVAs and the *post hoc* tests, respectively.

As expected, both indices decreased as a function of age. The main effects of age and all pairwise comparisons were

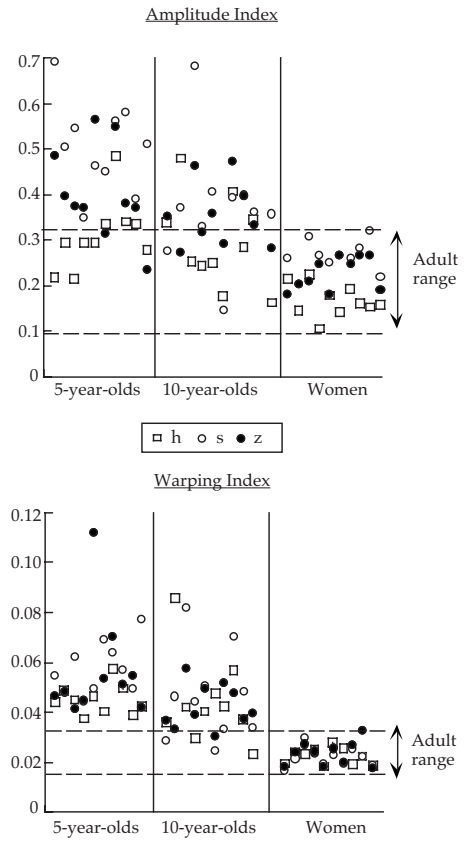


FIG. 3. Amplitude (top) and warping (bottom) indices for all speakers. Within the child groups, the leftmost five speakers are the girls; the rightmost are the boys. An open square represents /h/; an open circle represents /s/; a filled circle represents /z/.

significant for both measures (see Tables I and II). As shown in Fig. 3, however, the two indices have somewhat different developmental time courses. For the amplitude index, several values for both 5 and 10 year olds fall within the adult range (8 and 12 values out of 30, respectively). For the time warping index, relatively few of the 10 year olds' values (5 of 30) and *none* of the 5 year olds' values fell within the adult range. Thus, for this speech context, gestural magnitude appears, on the whole, to reach adultlike stability levels earlier than gestural phasing.

The main effect of consonant was significant for the amplitude index but not for the time warping index (see Table I). *Post hoc* tests showed that /h/ had lower amplitude variability than both /s/ and /z/ (see also Fig. 4). Pairwise comparisons for /s/ and /z/ were also significant. Figure 4 suggests that, qualitatively, the differences among the three consonants were the most extreme for the 5 year olds, but the age-by-consonant interaction was not significant presumably because of high within-group variability among the children combined with a modest sample size. Despite the mean differences between the three consonants (top panel of Fig. 4), it is evident from Fig. 3 (top panel) that not all speakers demonstrate the pattern of $h < z < s$ in their amplitude index. Individual patterns of variability are considered in Sec. III B.

To determine the relationships between the two indices, correlations were run for the whole data set (Fig. 5, top

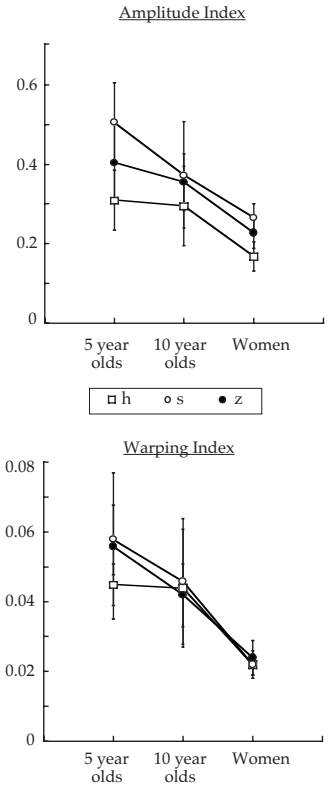


FIG. 4. Group averages for amplitude and warping indices for the three consonants. An open square represents /h/; an open circle represents /s/; a filled circle represents /z/.

panel) and then divided by consonant (middle) and age (bottom). For the combined data (top panel), the correlation was significant ($r=0.791$, $p<.0001$), with $r^2=0.625$. Thus, the two measures were moderately predictive of one another overall, but not totally redundant, since nearly 40% of the variance was unaccounted for. The middle panel of the figure shows that the correlation held about equally across the three consonants: All r-values were significant ($p<0.0001$), in the range of 0.76–0.86, and the slopes (m-values) were similar across /h s z/. When the data were split by age, however, clear group differences emerged (bottom panel). In particular, the two indices did not correlate for the women ($r=0.23$, $p=0.2235$), whereas they did for both groups of children. In adults, then, temporal and amplitude variabilities seem to be entirely independent for these speech sequences.

TABLE I. Results of ANOVAs.

	DFs	F	p
		Amplitude index	
Age	2	38.372	<0.0001
Cons	2	16.495	<0.0001
Age×Cons	4	1.486	0.2140
Residual	81		
		Warping index	
Age	2	49.014	<0.0001
Cons	2	1.188	0.3100
Age×Cons	4	1.155	0.3369
Residual	81		

TABLE II. Results of *post hoc* tests.

	Amplitude index p-value	Warping index p-value
Age effect		
Adults vs. 5 year olds	<0.0001	<0.0001
Adults vs. 10 year olds	<0.0001	<0.0001
5 year olds vs. 10 year olds	0.0031	0.0064
Consonant effect	p	p
/h/ vs. /s/	<0.0001	0.1426
/h/ vs. /z/	0.0014	0.2692
/s/ vs. /z/	0.0183	0.7140

B. Individual patterns of consonant-related variability

This section explores within-speaker patterns in amplitude and warping over time. This qualitative analysis supple-

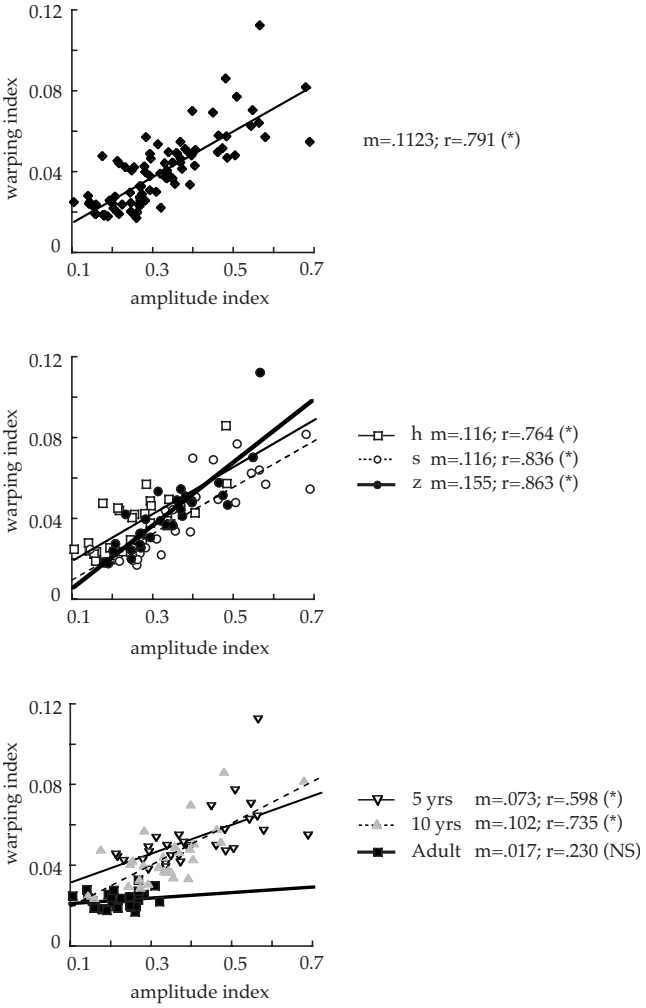


FIG. 5. Correlations between warping and amplitude indices for all data (top), the data split by consonant (middle), and the data split by speaker group (bottom). When split by consonant, open square, solid line represents /h/; open circle, dotted line represents /s/; and filled circle, heavy solid line represents /z/. When split by age group, open triangle, solid line represents 5 year olds; gray triangle, dotted line represents 10 year olds; and filled square, heavy line represents adults. Asterisks after the r-values indicate that the correlation was significant at $p<0.01$; NS indicates that the correlation was not significant. As noted for Fig. 2, the scales are unitless.

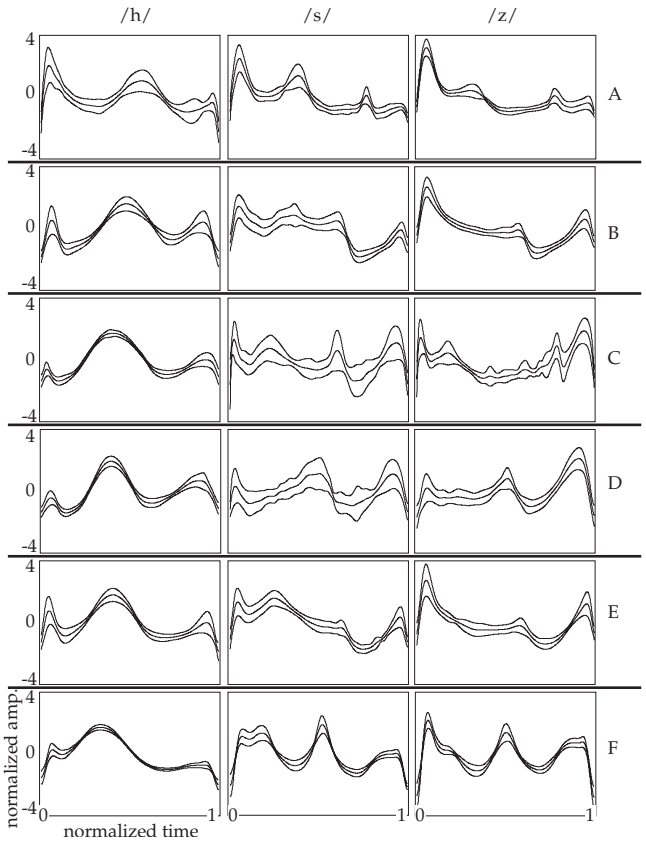


FIG. 6. Speaker differences in amplitude variability over time in /h/ (left), /s/ (middle), and /z/ (right). Top to bottom: row A, speaker 10F2 (10-year-old female). Row B, 5M1 (5-year-old male). Row C, 5F1 (5-year-old female). Row D, 10F3 (10-year-old female). Row E, 5M4 (5-year-old male). Row F, AF7 (adult female). Rows A and B show speakers who had relatively high variability in /h/ and /s/ but less in /z/; C and D represent speakers with relatively high variability in /s/ and /z/ but less in /h/; E and F show speakers where variability is similar across the three consonants.

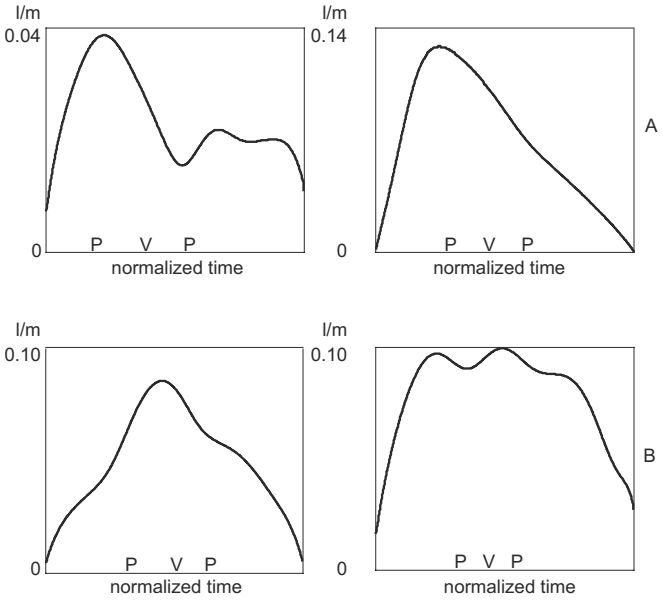


FIG. 7. Speaker differences in the location of maximum warping in /s/. The P V P notations in each figure show the temporal location of the two peaks and the valley characteristic of airflow during /s/. (A) Left: speaker AF7 (adult female); right: 10M3 (10-year-old male). (B) Left: speaker 5F6 (5-year-old female); right: 5M5 (5-year-old male). In all graphs, the *x*-axis is normalized time over the range of 0–1.

ments the quantitative index analysis above by providing a dynamic picture of amplitude and phasing variability over the VCV sequences. A key question is how variability over time compares among the three consonants.

Time-varying averages ±1 SD for a subset of speakers are given in Fig. 6. Recall that these average and SD functions were calculated on the aligned data. Thus, amplitude variability around the /h/ flow peak relates to differences in abduction degree after variations in temporal phasing have been factored out. Variability around the low-flow region of /z/ arises from differences in tongue-tip constriction degree. The typical profile for /s/ is an overall increase in airflow, resulting from vocal fold abduction, with a superimposed valley, representing the tongue-tip constriction (Klatt *et al.*, 1968). Of particular interest here is how variability during /s/ relates to that in /h/ and /z/. If both /s/ and /z/ show higher variability than /h/, it suggests that the speaker has less stable control over lingual constrictions than over laryngeal abduction. On the other hand, high variability for /s/ and /h/ but not for /z/ would imply more variable laryngeal control than supralaryngeal. Finally, it could be that variability in /s/ reflects supraglottal and glottal control to similar degrees. Figure 6 indicates that all three of these options may be observed across speakers, adults as well as children. For the two speakers shown in rows A and B, variability is high during /h/ (left panel) and /s/ (middle panel), whereas /z/ production (right panel) is more consistent. For the speakers in rows C and D, /s/ and /z/ have similar variability, whereas /h/ is less variable. Finally, the two speakers in rows E and F have similar variability across the three consonants; that is, /s/ appears to have about equal variability in laryngeal and supralaryngeal aspects of articulation.

Speakers also differed in terms of where their signals showed the most extreme warping. In the FDA warping algorithm, the edges of the records (that is, normalized times 0 and 1) are fixed in time, so phase variability is zero at those times. Between the beginnings and ends of the signals, however, considerable variation was observed, with no clear patterns based on age or consonant. Warping functions for a subset of speakers producing /s/ are given in Fig. 7. (Similar degrees of cross-subject variability were observed for the other two consonants.) The notations at the bottom of each graph indicate the relative temporal locations of the two peaks and the valley characteristic of /s/ production. For some speakers (row A), the most extreme warping was required early in the VCV during the increase in airflow that reflects abduction for /s/. In these data, it happened that no speakers had a clear "late" peak for /s/, but such cases were observed in /z/ and /h/. Most typical was for the warping peak to occur near the middle of the signal, here, around the time of maximum tongue-tip constriction for /s/, but speakers differed in whether there was a fairly short-term peak versus a more extended plateau (row B). In short, the dynamic patterns of amplitude and phasing variabilities revealed considerable cross-speaker differences.

IV. DISCUSSION

A. Temporal and spatial variability

1. Similarities and differences between amplitude and time warping measures

Past research on speech and other motoric behaviors has found higher token-to-token variability, on average, in children compared to adults. Many authors have appealed to aspects of neurological development to argue that children have more random sensorimotor noise than adults (Kent 1976, Lecours 1975, Netsell 1981, Tingley and Allen 1985, and Yan *et al.* 2000). At the same time, extensive differences have been observed within groups for specific variability measures and within individuals across measures (Smith, 1995, Smith and McLean-Muse, 1986, Stathopoulos, 1995, and Tingley and Allen, 1975), suggesting that variability in children may not be adequately characterized as simple random noise.

The use of FDA here allowed the production variability to be partitioned into two components. It was hypothesized that both amplitude and warping variabilities would be higher in children than in adults, but it was unclear whether the two indices would change in parallel over age. Significant group effects emerged for both measures. Thus, consistent with previous studies, children, as a group, were more variable than adults. Yet the indices also differed in two ways: (a) whereas there was considerable overlap between the 5 year olds' and women's amplitude indices, there was none at all for the warping index, and (b) a significant consonant effect was observed for the amplitude index but not for the warping index. The specifics of the amplitude and time warping index results are considered in more detail in Secs. IV A 2 and IV A 3. Here, the observation of primary interest is that the two measures do not show parallel developmental time courses.

Smith and Goffman (1998) also suggested that the temporal and spatial aspects of speech production might mature at different rates. As described in Sec. I B, their study measured lower-lip displacement over a sentence, carried out linear normalization, and then assessed (a) the relative timing of velocity peaks for three vocalic sequences and (b) the automatic classification success for waveform displacement in two consonant-vowel-consonant (CVC) syllables. Significant age differences were found for the relative timing measures but not for the classification success. There are many methodological differences between this work and that of Smith and Goffman (1998), including normalization techniques, ways of assessing temporal versus spatial variability, the duration of the speech sequences, and the data type (kinematics versus aerodynamics). Despite these differences, both studies found that temporal control in speech production, at least over sequences of a disyllable or longer, matures over a longer time period than amplitude variability or gestural shape.

These differences in temporal and spatial aspects of speech support the use of more sophisticated analyses of variability, such as those provided by FDA, to elucidate the nature of developmental changes across multiple dimensions. The finding that the two indices were not correlated in women further justifies considering them separately. Significant correlations in the children but not in the women could indicate greater overall noise in developing speakers for both temporal and amplitude domains. Yet such noise would not be truly random insofar as it is correlated across gestural magnitude and gestural phasing. While the current findings do not rule out a higher random noise factor in children, they do suggest that such a factor, alone, cannot explain all developmental differences.

As mentioned in Sec. I B, some developmental phonologists have also suggested that children may acquire articulatory gestures before they learn the appropriate phasing among them (Ferguson and Farwell, 1975; Vihman, 1993; Waterson, 1971). It must be noted that these observations were made on children much younger than those here and used a very different level of analysis, namely, phonetic transcription. It is not clear, therefore, whether the current results truly represent an example of the same phenomenon. More instrumental data, including studies of younger children, with varying speech sequence lengths and a wider variety of phonetic contrasts, are needed to ascertain whether the present findings are specific to these speech materials and methods or whether they indicate something more general about the nature of phonetic learning during childhood.

It is also unclear to what extent the current findings reflect purely motoric aspects of speech production versus ongoing linguistic development. Although children of the age studied here are usually considered to have a fairly complete sound inventory for their language, the details of phonological representations may well be refined over a longer time period. Further, speech differs from many tasks used in studies of nonspeech motor control in being a highly practiced skill. Again, assessing the generality of these results will require future studies that use similar analysis methods to explore a wider range of tasks, sampling across multiple articulators. Comparisons between speech perception and production abilities in the same children may also help to differentiate linguistic and motoric domains (see, e.g., Nittrouer and Studdert-Kennedy, 1987; Nittrouer *et al.*, 1989).

2. Amplitude index

The amplitude index in this study represents variability in airflow once phasing variability has been largely removed through nonlinear warping. The consonant effect for this measure indicates that airflow management is more consistent in /VhV/ sequences than in /VzV/ or /VsV/ for both adults and children. Since variability may increase with task demands (e.g., Yan *et al.*, 2000), these data provide some evidence that the lingual fricatives impose a greater motoric burden than the glottal fricative. Further, variability was highest, on average, for /s/, suggesting that the combination of abduction and supraglottal constriction, the most articulatorily complex sequence of the three, is also the most variable in gestural magnitudes. Although the warping index did not show significant consonant effects, an inspection of Fig. 4 (bottom panel) does suggest a possible qualitative difference in the 5 year olds, with less phasing variability in /h/

than in /s/ and /z/. More data might reveal that young children do not achieve equivalent phasing control of all articulators at the same time.

3. Time warping index

Many past speech studies have observed that individual children may be rather adultlike on some variability measures (Smith, 1995; Smith and McLean-Muse, 1986; Smith *et al.*, 1996; Stathopoulos, 1995). The 5 year olds' time warping data obtained here appear to represent an exception to that general conclusion: None of their 30 index values fell into the adult range.

One possibility that should be ruled out is that this finding results from some kind of processing artifact. Longer segment and syllable durations in children's speech compared to adults' (e.g., see Eguchi and Hirsh, 1969; Kent and Forner, 1980; Smith, 1978) are a potential source of artifact. Specifically, greater time warping for children's data could reflect a greater error "buildup" over a longer time sequence. Two considerations suggest that longer durations in children, all else being equal, should not bias them to higher warping indices.

First, as described in Sec. II E, putting the data into functional form involves expressing them in terms of a fixed number of spline functions. This effectively low-pass filters the data at a frequency that depends on the length of the record: longer records are filtered at lower frequencies. This should actually predispose signals with longer durations to have lower overall shape variability. (In this respect, FDA shares a drawback observed by Lucero *et al.* (1997) for one method of computing the STI.) Second, a short formal test on the effect of record length indicated that this factor, alone, does not yield higher amplitude or time warping indices. One woman's productions of /z/ were selected randomly from the data pool and manipulated as follows: The mean of the records was adopted as an initial pattern, and its length was scaled over a range of 0.5–5.0 of the original duration by linear interpolation. Random phase and amplitude variability components were then introduced into the signals. These components were generated from random time series [see Lucero (2005)] setting their SDs to 0.02 (for phase variability) and 0.2 (for amplitude), similar to the actual measured values for this speaker. Finally, the FDA algorithm was applied to this modified set of data, and warping and amplitude indices were obtained. The results, given in Fig. 8, indicate that simply increasing signal duration does not lead to higher variability indices but indeed, on the whole, to lower ones—especially for the time warping. More work is needed on establishing the precision of FDA computations as a function of signal parameters such as duration, but for present purposes it appears safe to conclude that higher time warping indices in children are not an artifact of longer durations but instead indicate real developmental differences.

B. Individual differences

Despite the general group tendencies revealed by the statistical results, the time-varying plots of mean /h s z/ ±1 SD in Fig. 6 indicate considerable variation in cross-consonantal patterns across speakers. Some participants had comparable variability in /s/ and /h/ (which have similar laryngeal requirements), whereas others had similar patterns in /s/ and /z/ (which have comparable tongue-tip activity). Further, the peak of the time warping SDs (Fig. 7) occurred in varying locations in the VCV sequence across speakers. Although these are qualitative assessments, they lend strength to the general conclusion that higher variability in children cannot be modeled solely as a higher noise factor that is added to all sound productions across the board. Within-speaker variability differed as a function of the articulatory task (i.e., consonant), and temporal variability was localized in the signal in speaker-specific locations. It appears that children adopt individual strategies for achieving speech goals, as has been observed for adults (Johnson *et al.*, 1993; Koenig *et al.*, 2005; Raphael and Bell-Berti, 1975).

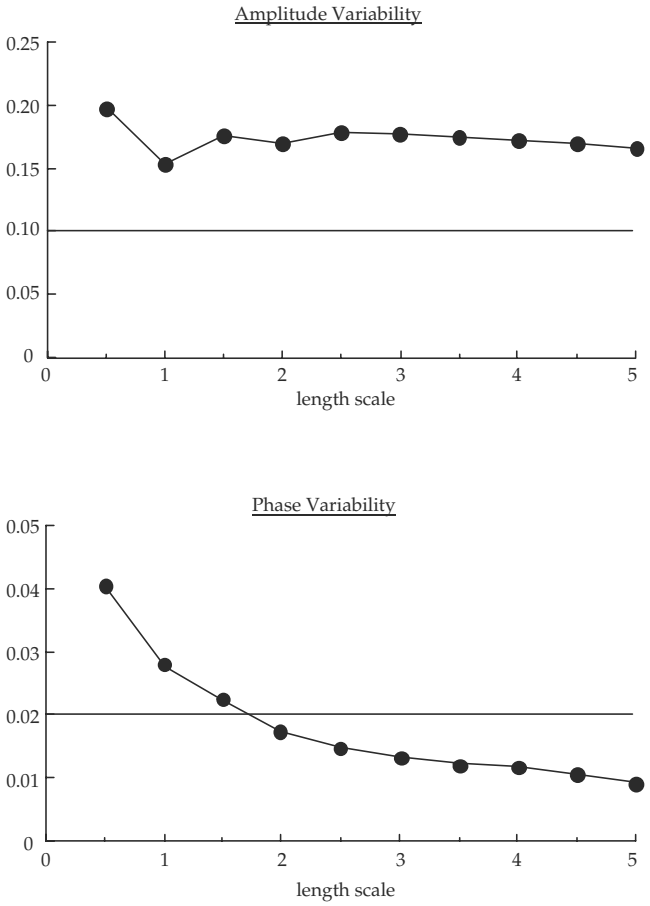


FIG. 8. Amplitude (top) and warping (bottom) indices as a function of increasing signal length.

V. CONCLUSIONS

Numerous studies have reported developmental differences in token-to-token variability. A common interpretation has been that random sensorimotor noise is higher in children than in adults, but many authors have drawn on individual differences or theoretical perspectives to emphasize individual learning and exploration. In this work, FDA was employed to decompose variability into two different components and to explore how variability was distributed over the time-normalized VCV sequences. Two findings argue

against attributing children's higher speech production variability to a single random noise factor: Temporal and amplitude variability did not display the same developmental time course, and considerable individual variation was observed in which consonants were the most variable and where in the disyllable the temporal variation was most extreme.

Perceptually adequate productions of /s/ and /z/ occur relatively late for some children. In the current data, /s/ and /z/ were more variable on average than /h/ for all age groups, and /s/ had the most amplitude variability of all, consistent with the prediction that the combination of laryngeal and supralaryngeal articulations for this sound presents an additional level of articulatory complexity. In spite of these group tendencies, some speakers showed different patterns. In particular, a comparison of variability across /h s z/ suggested that laryngeal actions make appreciable contributions to /s/ variability in some speakers.

Little work exists on how children learn to combine laryngeal and oral actions for voiceless fricatives. More data, across a wider range of ages and using a broader variety of speech tasks, are needed to determine the acquisitional time course for control of the larynx as compared to supralaryngeal articulators. Even if general trends emerge, however, the results here suggest that production consistency can differ widely across articulators in speaker-specific ways.

ACKNOWLEDGMENTS

Preliminary analyses and portions of this work were presented at the 140th and 143rd meetings of the Acoustical Society of America, the First Pan-American/Iberian meeting on Acoustics (Cancun), the 7th International Conference on Spoken Language Processing (Denver), and the 15th International Congress of Phonetic Sciences (Barcelona). We express our thanks to the following: (a) Linda Hunsaker, for preparing the stimulus pictures; (b) Elaine Russo Hitchcock, for assistance in subject recruitment and recording; (c) several students from Long Island University, Brooklyn and New York University, for assistance in recording and data analysis: in alphabetical order, Giridhar Athmakuri, Ingrid Davidowitz, Linda Greenwald, Karen Keung, and Gabrielle Rothman; (d) Elliot Saltzman, for raising the issue of possible artifact in longer signals; and (e) Julia Irwin, Anders Löfqvist, Martha Tyrone, Brad Story, and 3 anonymous reviewers for helpful comments on earlier versions of this paper. This work was supported by NIH Grant No. DC-04473-03 to Haskins Laboratories and by MCT/CNPq (Brazil).

[1]To our knowledge, the only other study that has examined variability using aerodynamic data is Zajac and Hackett (2002). Those authors assessed the timing of flow and pressure peaks in the /mp/ sequence of the word "hamper." Although aerodynamic signals were used, the aims and results of that study were more comparable to studies that have quantified temporal variability using acoustic signals than to the work presented here.

[2]Fewer than 15 productions were available in five cases (out of 90: 3 groups $\times$ 10 speakers $\times$ 3 consonants). All of these were in the 5 year olds, and three were from one child who produced 10–14 usable tokens each of /h s z/. Most of her /h/ exclusions reflected pausing; /s z/ exclusions were mostly due to voicing errors because she tended to perseverate when changing from one production to the next. Her data were not unusual in other respects; for example, her variability indices were not notably different from those of other 5 year olds.

[3]The number of splines was also selected by considering the expansion as a low-pass filtering process (Unser *et al.*, 1993a, 1993b). According to such theory, letting F be the sampling frequency of the tokens, n the number of samples in the tokens, N the number of spline functions and O their order, then the cutoff frequency f of the filtering process is $f=(F/2)^{\star}(N-O+2)/n$. In the present case, $N=42$, $O=4$, and $F=10$ kHz, and assuming $n=5000$ points (approximate average number of samples of the tokens), we obtain $f=40$ Hz. This frequency range is sufficient to yield a good representation of record shapes.

Berthier, N. E. (**1996**). "Learning to reach: A mathematical model," Dev. Psychol. **32**, 811–823.

Cheng, H. Y., Murdoch, B. E., Goozée, J. V., and Scott, D. (**2007**). "Electropalatographic assessment of tongue-to-palate contact patterns and variability in children, adolescents, and adults," J. Speech Lang. Hear. Res. **50**, 375–392.

Chermak, G. D., and Schneiderman, C. R. (**1986**). "Speech timing variability of children and adults," J. Phonetics **13**, 477–480.

Deutsch, K. M., and Newell, K. M. (**2001**). "Age differences in noise and variability of isometric force production," J. Exp. Child Psychol. **80**, 392–408.

Deutsch, K. M., and Newell, K. M. (**2003**). "Deterministic and stochastic processes in children's isometric force variability," Dev. Psychobiol. **43**, 335–345.

Deutsch, K. M., and Newell, K. M. (**2004**). "Changes in the structure of children's isometric force variability with practice," J. Exp. Child Psychol. **88**, 319–333.

Eguchi, S., and Hirsh, I. J. (**1969**). "Development of speech sounds in children," Acta Oto-Laryngol. **57**, 1–51.

Ferguson, C. A., and Farwell, C. B. (**1975**). "Words and sounds in early language acquisition," Language **51**, 419–439.

Goffman, L., Ertmer, D. J., and Erdle, C. (**2002**). "Changes in speech production in a child with a cochlear implant: Acoustic and kinematic evidence," J. Speech Lang. Hear. Res. **45**, 891–901.

Green, J. R., Moore, C. A., and Reilly, K. J. (**2002**). "The sequential development of jaw and lip control for speech," J. Speech Lang. Hear. Res. **45**, 66–79.

Hatze, H. (**1986**). "Motion variability: Its definition, quantification, and origin," J. Motor Behav. **18**, 5–16.

Howe, M. S., and McGowan, R. S. (**2005**). "Aeroacoustics of [s]," Proc. R. Soc. London, Ser. A **461**, 1005–1028.

Hughes, D., McGillivray, L., and Schmidek, M. (**1997**). *Guide to Narrative Language: Procedures for Assessment* (Thinking, Eau Claire, WI).

Jensen, J. L., Thelen, E., Ulrich, B. D., Schneider, K., and Zernicke, R. F. (**1995**). "Adaptive dynamics of the leg movement patterns of human infants: III. Age-related differences in limb control," J. Motor Behav. **27**, 366–374.

Johnson, K., Ladefoged, P., and Lindau, M. (**1993**). "Individual differences in vowel production," J. Acoust. Soc. Am. **94**, 701–714.

Kent, R. D. (**1976**). "Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies," J. Speech Hear. Res. **19**, 421–447.

Kent, R. D. (**1992**). "The biology of phonological development," in *Phonological Development: Models, Research, Implications*, edited by C. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 65–90.

Kent, R. D., and Forner, L. L. (**1980**). "Speech segment durations in sentence recitations by children and adults," J. Phonetics **8**, 157–168.

Kewley-Port, D., and Preston, M. (**1974**). "Early apical stop production: A voice onset time analysis," J. Phonetics **2**, 195–210.

Klatt, D. H., Stevens, K. N., and Mead, J. (**1968**). "Studies of articulatory activity and airflow during speech," in *Sound Production in Man*, edited by A. Bouhuys (New York Academy of Sciences, New York).

Koenig, L. L. (**2000**). "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds," J. Speech Lang. Hear. Res. **43**, 1211–1228.

Koenig, L. L., Mencl, W. E., and Lucero, J. C. (**2005**). "Multidimensional analyses of voicing offsets and onsets in female speakers," J. Acoust. Soc. Am. **118**, 2535–2550.

Körding, K. P., and Wolpert, D. M. (**2004**). "Bayesian intergration in sensorimotor learning," Nature (London) **427**, 244–247.

Lecours, A. R. (**1975**). "Myelogenetic correlates of the development of speech and language," in *Foundations of Language Development: A Mul-*

tidisciplinary Approach, edited by E. H. Lenneberg and E. Lenneberg (Academic, New York), Vol. **1**, pp. 121–135.
Lee, S., Byrd, D., and Krivokapic, J. (**2006**). "Functional data analysis of prosodic effects on articulatory timing," J. Acoust. Soc. Am. **119**, 1661–1671.
Lee, S., Potamianos, A., and Naryanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.
Lisker, L., Abramson, A. S., Cooper, F. S., and Schvey, M. H. (**1969**). "Transillumination of the larynx in running speech," J. Acoust. Soc. Am. **45**, 1544–1546.
Liu, Y.-T., Mayer-Kress, G., and Newell, K. M. (**2006**). "Qualitative and quantitative change in the dynamics of motor learning," J. Exp. Psychol. Hum. Percept. Perform. **32**, 380–393.
Lucero, J. C. (**1999**). "Computation of the harmonics-to-noise ratio of a voice signal using a functional data analysis algorithm," J. Sound Vib. **222**, 512–520.
Lucero, J. C. (**2005**). "Comparison of measures of variability of speech movement trajectories using synthetic records," J. Speech Lang. Hear. Res. **48**, 336–344.
Lucero, J. C., and Löfqvist, A. (**2005**). "Measures of articulatory variability in VCV sequences," ARLO **6**, 80–84.
Lucero, J. C., Munhall, K. G., Gracco, V. L., and Ramsay, J. O. (**1997**). "On the registration of time and the patterning of speech movements," J. Speech Lang. Hear. Res. **40**, 1111–1117.
Macken, M. A., and Barton, D. (**1980**). "The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants," J. Child Lang **7**, 41–74.
Miller, J. F., and Chapman, R. S. (**1981**). "Research note: The relation between age and mean length of utterance in morphemes," J. Speech Hear. Res. **24**, 154–161.
Munson, B. (**2004**). "Variability in /s/ production in chileren and adults: Evidence from dynamic measures of spectral mean," J. Speech Lang. Hear. Res. **47**, 58–69.
Netsell, R. (**1981**). "The acquisition of speech motor control: A perspective with directions for research," in *Language Behavior in Infancy and Early Childhood*, edited by Rachel E. Stark (Elsevier, New York), pp. 127–156.
Nittrouer, S. (**1993**). "The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study," J. Speech Hear. Res. **36**(5), 959–972.
Nittrouer, S., and Studdert-Kennedy, M. (**1987**). "The role of coarticulatory effects in the perception of fricatives by children and adults," J. Speech Hear. Res. **30**, 319–329.
Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (**1989**). "The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults," J. Speech Hear. Res. **32**, 120–132.
Ohala, J. (**1975**). "The temporal regulation of speech," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, London), pp. 431–453.
Ohde, R. N. (**1985**). "Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children," J. Acoust. Soc. Am. **78**, 1554–1561.
Ramsay, J. O., Munhall, K. G., Gracco, V. L., and Ostry, D. J. (**1996**). "Functional data analyses of lip motion," J. Acoust. Soc. Am. **99**, 3707–3717.
Ramsay, J. O., and Silverman, B. W. (**1997**). *Functional Data Analysis* (Springer-Verlag, New York).
Ramsay, J. O., and Silverman, B. W. (**2002**). *Applied Functional Data Analysis: Methods and Case Studies* (Springer-Verlag, New York).
Raphael, L. J., and Bell-Berti, F. (**1975**). "Tongue musculature and the feature of tension in English vowels," Phonetica **32**, 61–73.
Sander, E. K. (**1972**). "When are speech sounds learned?," J. Speech Hear Disord. **37**, 55–63.
Sawashima, M. (**1970**). "Glottal adjustments for English obstruents," Haskins Laboratories Status Report on Speech Research **21/22**, 187–199.
Schulz, G. M., Sulc, S., Leon, S., and Gilligan, G. (**2000**). "Speech motor learning in Parkinson Disease," J. Med. Speech-Lg. Path. **8**, 243–247.
Scully, C., Castelli, E., Brearley, E., and Shirt, M. (**1992**). "Analysis and simulation of a speaker's aerodynamic and acoustic patterns for fricatives," J. Phonetics **20**, 39–51.
Shadle, C. (**1990**). "Articulatory-acoustic relationships in fricative consonants," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht), pp. 187–209.
Sharkey, S. G., and Folkins, J. W. (**1985**). "Variability in lip and jaw movements in children and adults: Implications for the development for speech motor control," J. Speech Hear. Res. **28**, 8–15.
Smith, A., and Goffman, L. (**1998**). "Stability and patterning of speech movement sequence in children and adults," J. Speech Lang. Hear. Res. **41**, 18–30.
Smith, B. L. (**1978**). "Temporal aspects of English speech production: A developmental perspective," J. Phonetics **6**, 37–67.
Smith, B. L. (**1992**). "Relationships between duration and temporal variability in children's speech," J. Acoust. Soc. Am. **91**, 2165–2174.
Smith, B. L. (**1994**). "Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adults' speech," J. Phonetics **22**, 155–175.
Smith, B. L. (**1995**). "Variability of lip and jaw movements in the speech of children and adults," Phonetica **52**, 307–316.
Smith, B. L., Kenney, M. K., and Hussain, S. (**1996**). "A longitudinal investigation of duration and temporal variability in children's speech production," J. Acoust. Soc. Am. **99**, 2344–2349.
Smith, B. L., and McLean-Muse, A. (**1986**). "Articulatory movement characteristics of labial consonant productions by children and adults," J. Acoust. Soc. Am. **80**(5), 1321–1328.
Smith, B. L., Sugerman, M. D., and Long, S. H. (**1983**). "Experimental manipulation of speaking rate for studying temporal variability in children's speech," J. Acoust. Soc. Am. **74**, 744–749.
Sosnik, R., Hauptmann, B., Karni, A., and Flash, T. (**2004**). "When practice leads to coarticulation: The evolution of geometrically defined movement primitives," Exp. Brain Res. **156**, 422–438.
Stathopoulos, E. T. (**1995**). "Variability revisited: An acoustic, aerodynamic, and respiratory kinematic comparison of children and adults during speech," J. Phonetics **23**, 67–80.
Thelen, E., Corbetta, D., and Spencer, J. P. (**1996**). "Development of reaching during the first year: Role of movement speed," J. Exp. Psychol. Hum. Percept. Perform. **22**, 1059–1076.
Tingley, B. M., and Allen, G. D. (**1975**). "Development of speech timing control in children," Child Dev. **46**, 186–194.
Todorov, E., and Jordan, M. I. (**2002**). "Optimal feedback control as a theory of motor coordination," Nat. Neurosci. **5**, 1226–1235.
Unser, M., Aldroubi, A., and Eden, M. (**1993a**). "B-spline signal processing: Part I—Theory," IEEE Trans. Signal Process. **41**, 821–833.
Unser, M., Aldroubi, A., and Eden, M. (**1993b**). "B-spline signal processing: Part II—Efficient design and applications," IEEE Trans. Signal Process. **41**, 834–848.
Vetter, P., and Wolpert, D. M. (**2000**). "The CNS updates its context estimate in the absence of feedback," NeuroReport **11**, 3783–3786.
Vihman, M. M. (**1993**). "Vocal motor schemes, variation and the perception-production link," J. Phonetics **21**, 163–169.
von Hofsten, C. (**1989**). "Motor development as the development of systems: Comments on the Special Section," Dev. Psychol. **25**, 950–953.
Walsh, B., and Smith, A. (**2002**). "Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes," J. Speech Lang. Hear. Res. **45**, 1119–1133.
Waterson, N. (**1971**). "Child phonology: A prosodic view," J. Linguistics **7**, 179–211.
Yan, J. H., Stelmach, G. E., Thomas, J. R., and Thomas, K. T. (**2000**). "Developmental features of rapid aiming arm movements across the lifespan," J. Motor Behav. **32**, 121–140.
Zajac, D. J., and Hackett, A. (**2002**). "Temporal characteristics of aerodynamic segments in the speech of children and adults," Cleft Palate Craniofac J. **39**, 432–438.

Temporal organization of English clear and conversational speech[a)]

Rajka Smiljanić[b)] and Ann R. Bradlow[c)]
Department of Linguistics, Northwestern University, Evanston, Illinois 60208

(Received 26 April 2007; revised 31 July 2008; accepted 31 August 2008)

This study investigated the effect of hyperarticulated, intelligibility-enhancing clear speech on temporal characteristics as reflected in number, durations, and variability of consonant and vowel intervals in sentence- and paragraph-length utterances. The results of sentence-in-noise listening tests showed a consistent clear speech intelligibility gain across the utterances of varying complexity indicating that the talkers successfully maintained clear speech articulatory modifications throughout longer stretches of speech. The acoustic analysis revealed that some temporal restructuring accompanied changes in speaking style. This temporal restructuring was observed in the insertion of consonant and vowel segments that were dropped or coarticulated in conversational speech and in an increase in the number of prosodic phrases for clear speech. Importantly, coefficients of variation (variation of consonantal and vocalic intervals normalized for changes in speaking rate) for both consonantal and vowel intervals remained stable in the two speaking styles. Overall, these results suggest that increased intelligibility of clear speech may be attributed to prosodic structure enhancement (increased phrasing and enhanced segmentability) and stable global temporal properties. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2990712]

PACS number(s): 43.70.Fq, 43.71.Es, 43.71.Gv [CHS] Pages: 3171–3182

I. INTRODUCTION

As speech unfolds in time, more information about its temporal organization is revealed. From the talker's perspective, control over the temporal dimension at the segmental level involves merging multiple influences on speech planning. From the listener's perspective, a crucial feature of speech comprehension involves parsing the temporal dimension at the phonetic level into a hierarchically ordered, complex structure. Thus, from both points of view, the temporal organization of speech crucially involves the integration of short time-scale durations (segments) with longer time-scale temporal modulation patterns (words, phrases, and paragraphs). The goal of the present study was to look at the temporal structure of English as reflected in segment-level durations under conditions that induce global temporal changes, such as variation in speaking style and utterance complexity. Specifically, we investigated the effect of the listener-oriented speaking style called "clear speech" on temporal characteristics of segmental intervals in sentence- and paragraph-length utterances. This study forms part of a broad project whose overarching goal is to understand the signal-dependent and linguistic principles that guide the conversational-to-clear speech transformation in production and that may underlie the enhanced intelligibility of clear speech over conversational speech.

Clear speech is an intelligibility-enhancing speaking style that talkers naturally and spontaneously adopt when listeners have perceptual difficulty due to, for instance, a hearing loss or a different native language. Clear speech is related to other goal-oriented modes of speech production, such as infant-directed speech, Lombard speech, and computer-directed speech, in which talkers "adjust" their output to meet the demands of their target audience or the communicative situation (Junqua, 1993; Skowronski and Harris, 2005; Kuhl *et al.*, 1997; Fernald, 2000; Liu *et al.*, 2004). Clear speech also shares characteristics with variation due to prosodic strengthening and fast-to-slow rate modifications (Cole *et al.*, 2007; de Jong, 1995; Cho, 2005; Hirata, 2004; Miller and co-workers, 1988, 1989). These modifications can all be viewed as involving changes from hypo- to hyperarticulation (H&H Theory, Lindblom, 1990) with respect to articulatory and acoustic features such as reduction of target undershoot and enhancement of phonemic contrasts as seen in, for instance, expanded vowel spaces and increased vowel length and consonant voicing contrasts. Articulatory studies of prominence and of prosodic structure showed that stressed and accented syllables and segments at phrase boundaries were produced with larger, less overlapped gestures and with larger linguapalatal contact (de Jong, 1995; Byrd and Saltzman, 2003; Byrd *et al.*, 2000, 2006; Fougeron and Keating, 1997; Cho, 2006). In a more direct investigation of clear speech articulatory strategies, Perkell *et al.* (2002) found that consonant-volwel-consonant (CVC) syllables in clear speech were produced with an increased articulatory effort (greater peak speed and larger articulatory movement distance) beyond increased segment durations for some talkers (intelligibility was not measured). These studies, then, lend the support for linking acoustic clear speech findings with articulatory hyperarticulation. An important difference between hyperarticulations induced by

[a)]Portions of this work were presented at the 151st meeting of the Acoustical Society of America in Rhode Island (2006) and 10th Conference on Laboratory Phonology in Paris, France (2006)
[b)]Now in the Department of Linguistics, University of Texas at Austin. Electronic mail: rajka@austin.utexas.edu
[c)]Electronic mail: abradlow@northwestern.edu

prosodic structure and prominence and those due to clear speech changes may be that the former have been shown at the local (segment, syllable) level while clear speech may involve similar articulatory adjustments at a more global level, i.e., beyond stressed and accented syllables. Finally, while intelligibility of these hyperarticulated forms, such as stressed syllables and domain-initial segments, has not been investigated, conversational-to-clear speech modifications can provide us with a window into changes along the hypo- to hyperarticulation continuum with the overt purpose of enhancing intelligibility.

Previous work on English clear speech production and perception has established that naturally produced clear speech enhances intelligibility for various listener populations including adults with normal or impaired hearing (Picheny *et al.*, 1986; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Ferguson, 2004; Ferguson and Kewley-Port, 2002; Smiljanic and Bradlow, 2005), elderly adults (Schum, 1996; Helfer, 1998), both native and non-native listeners of the target language (Bradlow and Bent, 2002; Bradlow and Alexander, 2007; Smiljanic and Bradlow, 2007), and children with and without learning impairments (Bradlow *et al.*, 2003). Acoustic comparisons of conversational and clear speech in English have shown that clear speech modifications typically involve enhancement of the overall acoustic salience of the speech signal by means of a decreased speaking rate, longer and more frequent pauses, an expanded pitch range, greater sound pressure levels, more salient stop releases, greater obstruent intensity, increased energy in the 1000–3000 Hz range of long-term spectra, and increased modulation depth of low frequency modulations of the intensity envelope (Picheny *et al.*, 1986, 1989; Bradlow *et al.*, 2003; Liu *et al.*, 2004; Krause and Braida, 2004; Smiljanic and Bradlow, 2005). In addition, the vowel space is expanded in clear speech when compared to conversational speech (Picheny *et al.*, 1986; Moon and Lindblom, 1994; Johnson *et al.*, 1993; Ferguson and Kewley-Port, 2002, 2007). Vowel space expansion in clear speech was demonstrated in a cross-linguistic study by Smiljanic and Bradlow (2005) who showed equivalent vowel space expansion in English and Croatian despite the difference in vowel inventory size (more than ten in English but just five distinct vowel quality categories in Croatian). Thus, in addition to making the speech signal "stand out" more robustly from the masking effect of noise, clear speech production involves articulatory strategies that enhance vowel contrasts by making them spectrally more distinct from each other.

In the temporal domain, clear speech has been associated with segmental lengthening and an insertion of pauses and short segments, such as a schwa vowel at the ends of words or in voiced obstruent-semivowel clusters (Picheny *et al.*, 1986). Importantly, previous work has demonstrated nonuniform segmental lengthening. For instance, tense vowels were lengthened more than lax vowels in clear speech (Picheny *et al.*, 1986; Uchanski, 1988). Recently, in an in-depth, cross-language exploration of temporal changes for segmental contrasts in Croatian and English clear speech (relative to conversational speech), Smiljanic and Bradlow (2008) found similar asymmetrical patterns of lengthening for Croatian phonemically long and short vowels, English vowels preceding voiced and voiceless coda stops, and Croatian and English voice onset time (VOT) for voiced versus voiceless stop categories. However, the duration ratio between the "long" and "short" members of the contrastive categories in both English and Croatian was found to be remarkably stable across the two speaking styles suggesting that, unlike the spectral contrasts, duration contrasts are stable rather than enhanced under conditions of hyperarticulation for clear speech.

Similar to clear speech, nonuniform increases in segment durations have been found with changes in speaking rate (Miller *et al.*, 1986, 1988, 1989). A number of studies that examined the effect of speaking rate on temporal features of speech, such as VOT, short/long vowel duration, and single/geminate stop duration, argue for relational invariance in the production of these contrasts across speaking rates.[1] Although they found that the duration difference is enlarged between the two members of the contrasting pair when expressed in absolute measures, proportional measures exhibited stability across speaking rates for various languages (Hirata, 2004; Hirata and Whiton, 2005; Kessinger and Blumstein, 1998; Pickett *et al.*, 1999; Boucher, 2002). Some of these studies (e.g., Boucher, 2002) argue for the perceptual invariance as well as production invariance while others (Miller and Volaitis, 1989; Volaitis and Miller, 1992; Nagao and de Jong, 2007) show that the perceptual boundaries shift with changes in speaking rate. Importantly, all of these studies suggest that speakers and listeners rely on local timing relations when producing temporal contrasts and when judging category affiliation in a way that allows them to address changes in speaking rate.

In order to further investigate the nature of the temporal changes of clear speech relative to conversational speech, the present study examines durational characteristics more globally, i.e., we look at the vowel and consonant durations in terms of successive vowel and consonant intervals and their variability across whole utterances ranging from short sentences to a long paragraph (LP). This work builds on investigation of rhythmic characteristics of languages using similar measurements (Ramus *et al.*, 1999; Low *et al.*, 2000; Grabe and Low, 2002). Based on a set of duration measurements from the acoustic speech signal that presumably reflect the phonological properties of syllable structure and unstressed vowel reduction, these studies have made some progress toward linking impressionistically determined, typologically distinct rhythmic classes to durational characteristics of the speech signal. Stress-timed languages, such as English, tend to have a large number of syllable types as well as vowel reduction processes, which acoustically result in a smaller proportion of vocalic intervals within the sentence (%V) and greater variability in the vowel and consonant intervals ($\Delta V, \Delta C$) compared with syllable-timed languages, such as Spanish, which have fewer/simpler syllable types and no vowel reduction yielding higher %V and less variability in the intervocalic (consonantal) intervals.

The purpose of this study was to investigate in depth how the relative number, durations, and variability of successive segmental intervals (approximating closing and opening

gestures of the vocal tract) are affected by different speaking styles with an eye to linking these temporal changes to enhanced intelligibility. Moreover, we examined speech materials of various complexities ranging from isolated sentences to a lengthy and complex paragraph in order to observe temporal variability or stability under various discourse-level conditions. In our analyses, we avoid attributing segmental interval variability to any one feature of linguistic structure, such as overall rhythm class, prosodic structure, syllable complexity, phonological and inherent duration properties of segments, etc.; instead, we view segment durations as reflecting the conflation of multiple levels of linguistic structure into a single temporal dimension.

The specific aim of the present study was twofold. The first objective was to assess the ability of talkers to maintain conversational-to-clear speech articulatory modifications over the course of materials ranging from short sentences to a long and syntactically more complex paragraph (see Sec. II). The second objective of this study was to investigate whether/how conversational-to-clear speech modifications alter the durations of successive consonant and vowel intervals and their variability. In connection with our first goal, we were particularly interested in connecting previous clear speech work, which has focused mostly on intelligibility of isolated vowels, words, and short sentences (Picheny *et al.*, 1986; Gagne *et al.*, 2002; Bradlow and Bent, 2002; Ferguson and Kewley-Port, 2002; Krause and Braida, 2002; Smiljanic and Bradlow, 2005) to real world communicative situations that are more complex than those invoked in a laboratory and may require prolonged hyperarticulation on part of the talker. While the materials used in the present study, as in earlier studies, were read in a laboratory, they represent an attempt to move clear speech research toward more ecologically valid communication settings.

Our main prediction regarding the maintenance of clear speech across longer paragraphs is that, as with words and sentences, clear speech will be more intelligible than conversational speech. However, the overall intelligibility gain for clear speech paragraphs as a whole will not reveal whether enhanced intelligibility is present throughout the duration of the paragraph. With regard to this question, we envision three possible outcomes. Clear speech may require an effort on the part of the talker that is difficult to maintain during longer stretches of speech and as a result the clear speech intelligibility benefit may steadily decline throughout a paragraph. A second possibility is that the intelligibility gain for clear over conversational speech will remain quite constant across the paragraph. This would indicate a clear speech "resetting" that remains constant and adjusts to any changes in inherent intelligibility of the utterance due to the availability of contextual information, discourse and information structure, etc. A third possibility is that intelligibility could steadily increase toward the end of the paragraph indicating the combined beneficial effect of context and speaking style. In order to address the question of whether conversational-to-clear speech modifications and increased intelligibility can be maintained throughout the duration of a paragraph, we look at intelligibility scores for various portions of the paragraphs separately.

The second objective of this study was to investigate of whether/how durations of successive consonant and vowel intervals and their variability are affected by changes in speaking style). Previous research has established that slowing down in clear speech is realized as both the lengthening of segments and the insertion of more frequent pauses for monosyllabic words and short sentences (Picheny *et al.*, 1986; Krause and Braida, 2004; Ferguson and Kewley-Port, 2002; Bradlow *et al.*, 2003, Smiljanic and Bradlow, 2005, 2008). Furthermore, insertion of short segments was shown to characterize clear speech as well (Picheny *et al.*, 1986). We extend this previous work by providing a more global, typologically motivated set of measures including %C, %V, variability of C and V intervals, and coefficients of variability as well as phrasing for these intervals across utterances of various lengths and complexities. Based on the previous findings we expect that the durations of all segments will increase in clear speech. We also expect that the extent of variability of the short and long C and V intervals will change in clear speech compared to conversational speech. This could occur if, for instance, stressed syllables were lengthened more than unstressed syllables, thereby increasing the difference between short and long intervals. The increase in the interval variability would also be achieved through asymmetrical lengthening of various segmental duration contrasts as discussed above. Furthermore, a decrease in speaking rate and a reduction of coarticulation in clear speech could also result in the insertion of short C and/or V elements that were dropped or coarticulated with surrounding sounds in conversational speech (similar to the schwa insertion reported in Picheny *et al.*, 1986). These asymmetrical lengthening patterns along with the segmental insertion could lead to the improved intelligibility of clear speech through a more transparent reflection of prosodic structure and of underlying phonological structure (e.g., segment identification, word/syllable boundaries, and word-level stress patterns). Finally, it is possible that, although C and V variability increases in absolute numbers, proportional measures of this variability (normalizing for the changes in overall speaking rates) remain stable across two speaking styles (similar to the findings for the duration ratios for segmental contrasts found in Smiljanic and Bradlow, 2008).

This analysis presents a part of a larger effort to determine the consistent salient acoustic-phonetic correlates of increased speech intelligibility. However, it is important to note that a direct relationship between acoustic-phonetic variation due to changes in speaking style and variability in intelligibility has been difficult to establish (Uchanski *et al.*, 1996; Krause and Braida, 2002). Although numerous acoustic-phonetic features of conversational-to-clear style transformations have been identified, as noted above, it is not well understood yet if and how each of these modifications (including segmental lengthening and the overall temporal structure) affect intelligibility. Nevertheless, we hope to gain a better understanding of the acoustic-phonetic changes that talkers reliably and consistently produce in clear speech and that may underlie increased intelligibility. To this end, we provide a detailed analysis of temporal structure under various styles and discourse-level conditions.

II. METHODS

A. Participants

1. Production

Six native talkers of English (three female and three male) between the ages of 24 and 32 served as participants in the production study. They were graduate students in the Linguistics Department at Northwestern University and were all native speakers of general American English (GAE). None of the talkers had any known speech or hearing impairment at the time of recording. They were not aware of the purpose of the recordings. All participants were paid at the end of the recording session.

2. Perception

One hundred and twenty undergraduate students at Northwestern University, all native speakers of GAE, participated in sentence-in-noise perception tests. Half of the participants listened to the short sentences and half listened to the short and long paragraphs (SP and LP) (see Sec. II B). They received class credit for their participation in the listening test. Their ages ranged between 18 and 22 years. None of the listeners had any known speech or hearing impairment at the time of the test.

B. Stimuli

The stimuli consisted of three sets of materials: short sentences, one SP, and one LP. The 20 short sentences were taken from Ramus *et al.* (1999); they are short newslike declarative statements (e.g., *The next local elections will take place during the winter*). The mean number of syllables per sentence is 16 (range: 15–18). A SP, *The North Wind and the Sun*, is a standard text from phonetic research available in the Handbook of the International Phonetic Association (1999). Finally, a LP from contemporary literature (Sedaris, 2001) was included as well. The Sedaris paragraph is longer than the SP paragraph (212 versus 113 words) and has a more complex syntactic structure, including direct and reported speech, and, therefore, provides more opportunity for subjects to vary their prosody approximating spontaneous speech more closely than in the sentences or SP paragraph.

C. Procedure

1. Production

The subjects were recorded producing all three sets of materials in a sound-attenuated booth in the phonetics laboratory in the Department of Linguistics at Northwestern University. The participants first read the short sentences, which were written on index cards and randomized for each reading. Next, they read the SP followed by the LP. They read into a microphone and the speech was recorded directly to a disk at 24 bit accuracy using an Apogee PSX-100 analog-to-digital/digital-to-analog converter at a sampling rate of 16 kHz. Participants read all the materials first in conversational and then in clear speech. For the conversational style, the talkers were instructed to read as if they were talking to someone familiar with their voice and speech patterns. For the clear speaking style, the talkers were instructed to read as if they were talking to a listener with a hearing loss or a non-native speaker. The acoustic analyses of the recorded materials were done using PRAAT software for speech analysis (Boersma and Weenink, 2006).

2. Perception

After the recordings were made, the digital speech files were segmented into sentence-length files. For the two paragraphs, breaks were made at points of natural junctures (pauses, ends of phrases and sentences) and at points at which segmentation was possible. The North Wind and the Sun was segmented into 13 sentence-length files and the Sedaris (2001) paragraph was segmented into 31 files for each speaker. In order to obtain equivalent overall amplitude levels, all speech files were equated for rms amplitude[2] and then mixed with speech shaped noise at a −5 dB signal-to-noise ratio (SNR). The SNR level used in this experiment was chosen based on the results from previous studies with the goal of keeping the average intelligibility in the 45%–65% range (Smiljanic and Bradlow, 2005; 2008).

Each participant in the perception experiment heard either short sentences or the two paragraphs (SP and LP). In the short sentence condition, 60 listeners each heard 20 sentences (ten conversational and ten clear) produced by only one of the talkers (ten listeners per talker). Clear speech sentences always preceded conversational speech sentences so that any clear speech benefit obtained could not be explained by the subject's adaptation to the task or to the talker's speech patterns. For each talker, the ten sentences that were presented in the conversational style in one condition (in which half of the listeners participated) were presented in the clear speaking style in another condition (in which the other half of the listeners participated). This was done to guard against possible confounding effects of any particular sentence-style combination. The listeners never heard the same sentence twice.

Sixty new listeners participated in the paragraph condition. The clear speech paragraph always preceded the conversational style paragraph. Half of the listeners in the paragraph condition first heard the SP in clear speech followed by the LP in conversational speech. The other half of the listeners first heard the LP in clear speech followed by the SP in the conversational speaking style. Each listener heard both paragraphs produced by the same talker. In all conditions, each sentence/utterance in the listening test was preceded by a 400 ms leading silence and a 500 ms noise interval, and followed by a 500 ms noise interval.

The listeners were seated in front of a computer in a sound-attenuated booth in the phonetics laboratory in the Department of Linguistics at Northwestern University. Stimulus presentation was controlled by SUPERLAB PRO 2.01, a special-purpose experiment running software. Three practice sentences (from a different set of recordings by a different talker) were presented in both the short sentence and the paragraph conditions so that the subjects could get used to the nature of the stimuli mixed with noise and the procedure of advancing to the next trial. After each trial, the subject pressed the space bar on the keyboard to initiate the next trial. The listeners were instructed to write down every word

they heard. Each trial was presented only once but the duration of the pause between two trials was controlled by the subjects themselves.

D. Data analysis

1. Production

All the acoustic measurements were performed on the exact same sentences that were used in the sentence-in-noise perception tests. First, we marked consonant and vowel intervals for each sentence. Silences longer than 5 ms in duration excluding silent intervals preceding word-initial stop consonants were marked as pauses (following Smiljanic and Bradlow, 2005). All obstruents and sonorants that could be segmented out from the surrounding sounds were marked as belonging to consonantal intervals. If clear formant changes and amplitude differences were observed for glide-vowel sequences, for instance, glides were excluded from the vocalic portions. If no observable boundary could be found, the glides were included in the vocalic portion. All vowels were marked as vocalic intervals. In general, standard segmentation criteria were followed (Peterson and Lehiste, 1960; Lehiste, 1970). A contiguous sequence of one or more vowels (consonants) was marked as one vocalic (consonantal) interval even when it spanned syllable and word boundaries. If it was not possible to segment a consonant from the surrounding vocalic segments due to extensive coarticulation and lack of clear acoustic cues marking its beginning and end it was marked as a part of a vocalic interval. In this respect sonorants were more likely to be marked as vocalic although some obstruents were also highly coarticulated/weakened or completely dropped in productions. Such "segmentability" difficulties occurred more often in conversational speech than in clear speech (see Sec. III). Our segmentation process was, thus, based on acoustic rather than phonological criteria and in this respect resembles closely those of Ramus *et al.* (1999) and Grabe and Low (2002).

Once consonant and vowel intervals were labeled, the total number of consonant and vowel intervals and their durations were measured for each speaking style. For all measurements, subsequent calculations were done within prosodic phrases (PPs) and then averaged across PPs for different materials, speaking styles, and individual talkers. PPs were defined as stretches of speech delimited by pauses. We consider PPs to reflect phrasing implemented by talkers to emphasize the discourse structure of produced speech. In short sentence productions, PPs were mostly identical to the sentences themselves. In longer paragraphs, PPs sometimes corresponded with syntactic sentences and sometimes with intonation phrases (IPs) as defined by ToBI, a convention for transcribing the intonation and prosodic structure of spoken utterances (Silverman *et al.*, 1992). Some PPs encompassed several IPs.

The overall speech rate (vocalic intervals per second) was calculated by dividing the number of vocalic intervals in a PP by the total duration of that PP. Because we wanted to examine segmental lengthening independent of pauses in clear speech, pause duration was excluded from this calculation. Speaking rate was determined by the number of vocalic intervals measured rather than by the phonological number of syllables for each sentence. Although this measure may underestimate the number of phonologically defined syllables produced (due to the conflation of VV sequences into one V interval), it captures the variability in the articulation of syllables/segments (rate of syllable omission and the extent of coarticulation) as produced across individuals and speaking styles. The overall speaking rate for each speaking style and type of material was then determined by averaging the vocalic interval per second rate over all PPs produced by each talker.

Proportion of vowel and consonant intervals (%V,%C) and their variability (ΔV,ΔC) were also calculated %C and %V were calculated by dividing the total duration of C and V intervals, respectively, within a PP with the total duration of that PP. ΔV and ΔC are standard deviations of all vocalic and all consonant intervals within a PP. Finally, in order to take into account the overall slowing down of speech rate for clear versus conversational speech, we calculated coefficients of variation for vowel and consonant intervals as the standard deviation of the consonant or vowel intervals divided by the average duration of the consonant or vowel intervals for each PP. To help us compare variability in overall consonant and vowel interval duration for clear versus conversational speech without regard for the source of that variability (e.g., stress, vowel reduction, final lengthening, or inherent vowel duration), all vowel and consonant intervals within a PP were included in these calculations.

2. Perception

Each participant in the short sentence-in-noise perception test received a keyword correct score out of 107 for the 20 sentences they heard. Each participant in the paragraph-in-noise perception test received a keyword correct score out of 50 for the 13 sentences they heard in the SP and out of 109 for the 31 sentences they heard in the LP. For all materials, all content words counted as keywords. A keyword was counted as correct only if all morphemes of the target word were present and transcribed correctly, e.g., if the target word was "keeping," "keep," "keeps," or "kept" were scored as incorrect. Percentage correct scores were calculated and then converted to rationalized arcsine transform units (RAU) for statistical analyses (Studebaker, 1985). The transformed scores were then coded as RAU scores for −5 dB SNR conversational style and for −5 dB SNR clear style for each talker in the sentence and paragraph conditions.

III. RESULTS

A. Perception

The average perception scores (percentage correct keywords) and the average intelligibility gain (clear-conversational scores) for the short sentences, SP, and LP for each talker are given in Table I. The talkers are ordered by the amount of conversational-to-clear speech intelligibility gain in the short sentence condition from smallest to largest. Letters F or M designate a female or a male talker. The numbers refer to the recording order.

TABLE I. Average perception scores (percentage correct keywords) in conversational and clear speech and the average intelligibility gain (clear-conversational scores) for the short sentences, short paragraph (SP), and long paragraph (LP) for each talker.

	Intelligibility (%)								
	Short sentences			SP			LP		
Talker	Conv.	Clear	Clear-conv.	Conv.	Clear	Clear-conv.	Conv.	Clear	Clear-conv.
F6	47	51	4	34	35	2	32	44	12
F2	37	42	5	18	29	10	37	41	4
M1	38	53	15	26	25	−2	30	52	22
M5	40	62	22	44	84	40	45	71	26
M4	55	81	26	34	57	22	45	62	17
F3	39	74	35	21	60	39	39	53	14
Average	43	60	18	30	49	19	38	54	16

A repeated-measures analysis of variance (ANOVA) with material type (short sentence versus SP versus LP) and style (conversational versus clear) as within-subjects factors was performed on the RAU transformed percent correct scores (averaged for each listener in each style and material condition). There was a significant main effect of style on intelligibility score: $F(1,5)=15.745$, $p<0.05$. The effect of material type did not reach significance: $F(1,5)=5.034$, $p=0.075$. The two-way style by material type interaction was not significant either: $F(1,5)=0.085$, $p=0.783$. The results of the statistical analysis showed that intelligibility was not significantly different for short sentences and SP and LP, and, furthermore, that the clear speech intelligibility increase was equivalent for the three types of materials. Most talkers' conversational-to-clear speech modifications resulted in a moderate to large increase in intelligibility gain (range: 5%–40%). Intelligibility increase of less than 5% in clear speech was found for F6 for short sentences and SP, for F2 for LP and for M1 for SP.

In order to address the question of whether conversational-to-clear speech modifications and increased intelligibility are maintained throughout the duration of a paragraph, we divided each paragraph into four portions (0%–25%, 25%–50%, 50%–75% and 75%–100%) based on the total number of target keywords for each type of material. The average intelligibility scores (percentage keyword correct) for the two speaking styles in the four portions of each paragraph are given for each talker separately in Table II.

For most talkers in both paragraphs, clear speech was more intelligible than conversational speech in all four portions suggesting that talkers successfully maintained clear speech articulatory modifications throughout the duration of the paragraphs. For both paragraphs, listeners tended to have most difficulty in correctly identifying words in the initial portion (0%–25%). There was an increase in intelligibility after the initial portion (25%–50%), a slight decrease in the third quarter of the paragraph (50%–75%), and another increase during the last portion (75%–100%).

A three-way repeated-measures ANOVA with style (conversational versus clear), paragraph (SP versus LP), and portion (1, 2, 3, and 4) as within-subjects factors showed a significant main effect of style: $F(1,5)=13.616$, $p<0.05$. The portion factor was also significant: $F(3,15)=43.935$, $p<0.001$. However, the effect of the paragraph was not significant: $F(1,5)=4.939$, $p=0.077$. Only the paragraph by portion two-way interaction was significant: $F(3,15)=15.308$, $p<0.001$. The three-way interaction was not significant. Separate repeated-measures ANOVAs for the two paragraphs showed a main effect of portion [$F(3,15)=32.568$, $p<0.01$] but not of style [$F(1,5)=5.694$, $p=0.063$] only for SP. The two-way style by portion interaction was also significant: $F(3,15)=3.316$, $p=0.049$. Paired comparisons revealed that the effect of style was significant only for portions 2 and 3: $t(5)=2.492$, $p=0.05$; $t(5)=3.187$, $p<0.05$. Intelligibility in portions 1 and 4 was not significantly different for the two speaking styles: $t(5)=1.566$, $p=0.178$; $t(5)=1.77$, $p=0.137$. In contrast, both style and portion factors were significant for LP: [style: $F(1,5)=22.616$, $p<0.01$; portion: $F(3,15)=10.731$, $p<0.01$]. The two-way interaction was not significant: $F(3,15)=0.578$, $p=0.639$. For both paragraphs, then, overall intelligibility varied across the four portions. The clear speech effect was rather consistent throughout for the LP while it was not present in all portions for the SP.

Adding to the large body of work on factors that condition speech intelligibility, the results of the present study

TABLE II. Average intelligibility scores (percentage keyword correct) for conversational and clear speech in four portions of the short paragraph (SP) and long paragraph (LP) for each talker. Each score is averaged across listeners ($n=60$).

Paragraph	Style	Talker	25%	50%	75%	100%
SP	Conv.	F6	17	37	35	45
		F2	7	32	13	20
		M1	13	34	17	42
		M5	27	55	25	66
		M4	25	40	27	46
		F3	2	20	12	49
		Average	15	36	21	45
SP	Clear	F6	20	40	38	42
		F2	7	35	38	34
		M1	3	40	23	32
		M5	75	95	77	88
		M4	47	69	48	62
		F3	42	74	62	62
		Average	32	59	48	53
LP	Conv.	F6	29	33	28	39
		F2	41	37	40	35
		M1	19	33	27	40
		M5	45	59	37	41
		M4	40	53	43	43
		F3	36	45	41	37
		Average	35	43	36	39
LP	Clear	F6	43	53	40	42
		F2	26	44	50	45
		M1	47	59	55	49
		M5	52	84	70	75
		M4	58	70	53	68
		F3	50	57	45	61
		Average	46	61	52	57

pointed to a significant contribution of utterance-specific characteristics and speaking style to overall intelligibility. Portion-by-portion results revealed that intelligibility varied across paragraph-length utterances in both speaking styles (higher intelligibility in second and fourth portion compared to first and third portions) suggesting that context, choice of lexical items, introduction of a new topic, etc., all contribute to intelligibility significantly. Furthermore, different intelligibility results for the two paragraphs indicated that the nature of the text read may also affect speaking effort and accuracy, which in turn influence listening accuracy (SP is a familiar short fable while LP is a new unfamiliar conversationlike text, which may have caused a more animated/ hyperarticulated reading style). Additionally, LP could be easier perceptually than SP due to its lexical context and the nature of the material. Importantly, the intelligibility results illustrated that some of the perceptual difficulties that listeners face can be overcome by naturally produced clear speech in both sentence- and paragraph-length utterances. It is worth noting, though, that clear speech did not wipe out the detrimental effects of other factors, such as lack of contextual cues, lexical choices, or articulatory precision that varied within the course of the paragraphs. However, these factors will have to be addressed in future work, because the present materials were not designed to investigate them in a controlled manner. With regard to our initial question of whether clear speech can be maintained in paragraph-length utterances, these results demonstrated that clear speech can be maintained throughout longer stretches of speech. This expands on previous clear speech findings, which focused mostly on intelligibility of isolated vowels/syllables, words, and short sentences (Picheny *et al.*, 1986; Gagne *et al.*, 2002; Bradlow and Bent 2002; Ferguson and Kewley-Port, 2002; Krause and Braida, 2002; Smiljanic and Bradlow, 2005).

B. Production

The average results of the various duration measurements for each material in each speaking style for each talker are given in Table III. The results of the repeated-measures ANOVAs with style (conversational versus clear) and material type (sentence versus SP versus LP) as within-subjects factors on various dependent duration measures given in Table III.

C lengthening in clear speech was statistically significant: $[F(1,5)=26.118,\ p<0.01]$. The effect of the material type on C lengthening was significant as well: $[F(2,10)=19.091,\ p<0.001]$. The style by material type interaction was not significant: $[F(2,10)=1.6,\ p=0.250]$. V lengthening in clear speech was statistically significant as well: $[F(1,5)=13.412,\ p<0.05]$. The effect of the material type on V lengthening was not significant: $[F(2,10)=2.241,\ p=0.157]$. The style by material type interaction was not significant: $[F(2,10)=0.593,\ p=0.157]$. These results showed that talkers lengthened both C and V intervals significantly in clear speech. The effect of material type on C lengthening reflects a difference in the segmental makeup of the test materials.

Short sentences were typically produced as one phrase and were not conducive for pause insertion in either speaking style due to their length and relative syntactic simplicity. In contrast, syntactically more complex paragraphs showed an increase in the number of pauses in clear speech. More pauses, furthermore, indicated increased phrasing on average by three and eight phrases in SP and LP, respectively.

Speaking style had a significant effect on speaking rate: $[F(1,5)=43.5,\ p<0.001]$. The effect of the material type on speaking rate was significant as well: $[F(2,10)=7.263,\ p<0.05]$. The style by material type interaction was not significant: $[F(2,10)=0.035,\ p=0.121]$. The results showed that speaking rate was significantly decreased for all material types in clear speech. Short sentences spoken in the conversational speaking style had the fastest speaking rate while SP spoken in clear speech had the slowest speaking rate. The largest clear speech speaking rate decrease occurred for short sentences. These results suggest that the complexity and length of the planned speech affect the speaking rate. Furthermore, the amount of a speaking rate decrease in clear speech may be dependent on the baseline conversational speaking rate.

Next, we examined whether the proportional relation between C and V intervals changed in clear speech for short sentences and the two paragraphs. This measurement allowed us to examine whether segmental lengthening associated with a clear speech decrease in speaking rate affected C and V intervals equally or whether consonants and vowels were lengthened asymmetrically in clear speech. Since %C and %V are directly related (as one increases the other decreases) we discuss %C only. The results showed that consonantal intervals take up slightly more than half of the PP duration in both speaking styles. A repeated-measures ANOVA revealed that only material type had a significant effect on %C: $F(2,10)=16.076,\ p<0.05$. The effect of style was not significant: $[F(1,5)=0.486,\ p=0.157]$. The style by material interaction was not significant either: $[F(2,10)=0.020,\ p=0.981]$. These results showed that the average %C and %V differed for the three types of materials suggesting that syllable structure varied for short sentences, SP and LP with %C being highest for the SP. This is not surprising given that SP was constructed so that it contains all sounds of English as well as a wide variety of syllable types. The overall relationship between %C and %V, however, remained stable across two speaking styles for all material types, i.e., clear speech lengthening overall affected C and V intervals equally (%C is decreased in clear speech by only 0.42, 0.30, and 0.29 percentage points for short sentences, SP, and LP, respectively).

Even though the proportion of C and V intervals remained stable in the two speaking styles, the extent of variability of the C and V intervals may change in clear speech compared to conversational speech. The average variability of C and V intervals (standard deviations) is given in Table III. A repeated-measures ANOVA showed a significant main effect of only style on C variability: $F(1,5)=15.057,\ p<0.05$. The effect of material type was not significant: $[F(2,10)=0.797,\ p=0.477]$. The style by material interaction was not significant either: $[F(2,10)=2.647,\ p=0.0119]$. There was also a significant main effect of style:

TABLE III. Various C and V interval duration measurements in conversational and clear speech for short sentences and short paragraph (SP) and long paragraph (LP) for each talker: total number of C and V intervals and pauses, average speaking rate (vocalic interval/s), average C and V interval duration (s), average proportion of C and V intervals, average variability of C and V interval duration (standard deviations), and average coefficient of variance of C and V intervals. The averages were first calculated within PPs and then across PPs for each material type and individual talkers.

Material	Style	Talker	Total No. of C	Total No. of V	Total No. of pause	Avg. spkg. rate (voc. inters./s)	Avg. C dur. (s)	Avg. V dur. (s)	%C	%V	stdev C (s)	stdev V (s)	Coeff. of variance C	Coeff. of variance V
Short sent.	Conv.	F6	285	276	1	5.46	0.10	0.09	57.78	42.22	0.05	0.04	0.56	0.55
		F2	297	289	1	5.45	0.11	0.08	58.22	41.78	0.06	0.05	0.53	0.57
		M1	268	253	1	5.61	0.10	0.07	58.29	41.71	0.06	0.04	0.53	0.60
		M5	299	285	8	5.35	0.10	0.09	53.24	46.76	0.07	0.05	0.58	0.56
		M4	273	268	0	4.77	0.12	0.09	57.58	42.42	0.06	0.05	0.58	0.59
		F3	301	292	0	5.36	0.10	0.09	52.64	47.36	0.06	0.06	0.60	0.61
		Avg.	287.1	276.6	1.83	5.33	0.10	0.09	56.29	43.71	0.06	0.05	0.56	0.58
	Clear	F6	309	297	0	4.89	0.11	0.08	54.54	45.46	0.06	0.05	0.56	0.57
		F2	308	297	0	4.53	0.13	0.09	58.77	41.23	0.07	0.05	0.57	0.58
		M1	299	282	6	5.12	0.11	0.08	57.21	42.79	0.06	0.05	0.51	0.61
		M5	334	303	0	4.84	0.11	0.10	53.17	46.83	0.06	0.05	0.57	0.53
		M4	302	293	0	3.45	0.15	0.12	58.32	41.68	0.10	0.07	0.62	0.54
		F3	306	295	0	4.13	0.13	0.11	53.19	46.81	0.08	0.07	0.60	0.59
		Avg.	309.6	294.5	1.00	4.49	0.12	0.10	55.87	44.13	0.07	0.06	0.57	0.57
SP	Conv.	F6	136	131	7	5.15	0.11	0.08	58.06	41.94	0.06	0.04	0.56	0.55
		F2	134	131	7	4.75	0.12	0.08	58.96	41.04	0.07	0.05	0.54	0.53
		M1	135	130	7	5.47	0.10	0.07	59.76	40.24	0.05	0.04	0.51	0.58
		M5	135	129	8	5.29	0.11	0.08	58.74	41.26	0.05	0.04	0.51	0.50
		M4	133	124	13	4.04	0.14	0.10	57.06	42.94	0.07	0.06	0.51	0.58
		F3	135	129	7	5.09	0.11	0.08	56.98	43.02	0.06	0.04	0.55	0.53
		Avg.	134.6	129.0	8.17	4.97	0.11	0.08	58.26	41.74	0.06	0.05	0.53	0.55
	Clear	F6	139	134	9	4.98	0.11	0.09	56.55	43.45	0.06	0.05	0.57	0.53
		F2	139	133	12	3.92	0.15	0.10	59.94	40.06	0.08	0.06	0.53	0.57
		M1	134	127	10	4.90	0.12	0.08	58.71	41.29	0.06	0.05	0.52	0.59
		M5	139	135	9	4.20	0.13	0.10	57.60	42.40	0.07	0.05	0.55	0.51
		M4	146	137	14	3.65	0.15	0.11	57.69	42.31	0.09	0.05	0.58	0.45
		F3	138	132	12	3.85	0.14	0.11	57.31	42.69	0.08	0.06	0.59	0.54
		Avg.	139.1	133.0	11.00	4.25	0.13	0.10	57.96	42.04	0.07	0.05	0.56	0.53
LP	Conv.	F6	315	298	28	5.05	0.10	0.09	55.07	44.93	0.06	0.05	0.55	0.58
		F2	315	295	27	4.81	0.11	0.09	58.17	41.83	0.06	0.06	0.55	0.63
		M1	279	266	21	5.22	0.10	0.08	57.78	42.22	0.06	0.05	0.54	0.61
		M5	306	283	30	5.06	0.10	0.08	58.22	41.78	0.07	0.04	0.60	0.51
		M4	324	289	36	4.46	0.11	0.09	57.34	42.66	0.07	0.06	0.63	0.56
		F3	297	285	16	5.14	0.10	0.09	51.11	48.89	0.05	0.06	0.57	0.61
		Avg.	306.0	286.0	26.33	4.96	0.10	0.09	56.28	43.72	0.06	0.05	0.57	0.58
	Clear	F6	324	301	32	4.74	0.10	0.09	53.74	46.26	0.06	0.05	0.57	0.53
		F2	320	302	33	4.28	0.13	0.10	58.94	41.06	0.07	0.06	0.55	0.60
		M1	313	295	26	4.90	0.11	0.08	59.42	40.58	0.06	0.05	0.54	0.60
		M5	321	302	37	4.59	0.11	0.10	55.33	44.67	0.07	0.05	0.56	0.51
		M4	343	302	49	3.86	0.13	0.12	55.28	44.72	0.07	0.07	0.57	0.54
		F3	313	297	28	4.40	0.11	0.11	53.25	46.75	0.06	0.07	0.54	0.61
		Avg.	322.3	299.8	34.17	4.46	0.12	0.10	55.99	44.01	0.07	0.06	0.55	0.56

$F(1,5)=25.250$, $p<0.01$ and of material type: $F(2,10)=6.205$, $p<0.05$ on V variability. The style by material interaction was not significant: $[F(2,10)=0.736,\ p=0.503]$. The results revealed a significant increase in variability of C and V intervals for all three types of materials in clear speech (C variability increases in clear speech were 0.012, 0.013, and 0.005 s and V variability increases in clear speech were 0.009, 0.006, and 0.004 s for short sentences, SP, and LP, respectively). The overall V variability differed significantly for the three types of materials (lowest for SP and highest for LP) due to the varied syllabic complexity across short sentences and the two paragraphs. This was already observed in the difference in %C and %V for the three different materials.

Next, we examine the number of V and C intervals in conversational and clear speech as well as the coefficients of C and V interval variation (normalizing for changes in speaking rate).

ANOVA results for the number of C intervals showed a main effect of material type [$F(2,10)=600.342$, $p<0.001$] and of style [$F(1,5)=31.66$, $p<0.01$]. The two-way style by material type interaction was also significant: $F(2,10)=6.885$, $p<0.05$. Similar results were obtained for the number of V intervals: style $F(1,5)=26.295$, $p<0.01$; material type $F(2,10)=1777.389$, $p<0.001$; style by material type interaction $F(2,10)=152.028$, $p<0.05$. As expected, these results showed that the total number of C and V intervals varied for the short sentences, SP, and LP due to the difference in their length and complexity. Importantly, the results revealed a significant increase in the number of C: [$F(1,5)=31.66$, $p<0.01$] and V: [$F(1,5)=26.295$, $p<0.01$] intervals in clear speech compared with conversational speech. The number of C intervals was increased by 22, 4.5, and 16.33 and the number of V intervals was increased by 17.83, 4, and 13.83 for short sentences, SP, and LP, respectively.

ANOVA results revealed a significant effect of material type on the number of C intervals: [$F(2,10)=600.342$, $p<0.001$] and V intervals: [$F(2,10)=1777.389$, $p<0.001$]. The style by material interaction was also significant for the number of C intervals: [$F(2,10)=6.885$, $p<0.05$] and V intervals: [$F(2,10)=152.028$, $p<0.05$]. The significant two-way style by type of material interaction on the number of C and V intervals revealed that clear speech C and V increase differed for short sentences, SP, and LP. The smallest increase in the number of intervals was found for the SP. This paragraph was overall the shortest and provided the fewest opportunities for dropping of segments/syllables in conversational speech. The largest clear speech increase for both C and V intervals was found in short sentences. The fastest speaking rate (which occurred in the production of short sentences compared to SP and LP) induced the most dropping and coarticulation of segments, while the largest speaking rate decrease in clear speech (which also occurred for the short sentences) increased the number of inserted segments/syllables. Note that most previous work examined segment durations (speaking rate) and insertions in short sentences (Picheny *et al.*, 1986, Krause and Braida, 2002; Smiljanic and Bradlow, 2008). The present results indicate that some of the conversational-to-clear speech changes reported for short sentences in previous work may overestimate the presence of these features in longer utterances. It is, therefore, crucial to compare the production of short sentences with longer and more complex materials and their interaction with changes in speaking rate in an attempt to better understand speech communication in real-life situations. The results reported here also suggest that talkers may implement varied clear speech changes for different materials and different communicative demands.

In order to account for the differences in speaking rates (and the difference in the average C and V durations) between the two speaking styles, we examined the coefficient of variation for the two intervals. The coefficient of variation was calculated as a standard deviation of the C or V intervals divided by the average duration of the C or V intervals for each PP. The average results across PPs for each talker are given in Table III. ANOVA results revealed that style and type of material did not have a significant effect on C coefficient of variation: style: [$F(1,5)=2.72$, $p=0.160$]; material: [$F(2,10)=3.186$, $p=0.085$]. The two-way interaction was not significant either: [$F(2,10)=3.328$, $p=0.078$]. There was a significant main effect of material type on V coefficient of variation: $F(2,10)=8.527$, $p<0.01$, but not of style: [$F(1,5)=1.943$, $p=0.222$]. The two-way interaction was not significant: [$F(2,10)=0.040$, $p=0.961$]. The results showed that the coefficient of variation differed for the three types of the materials (lowest for the SP and similar for the short sentences and the LP). However, the results for the coefficient of variation for C and V intervals showed that when speaking rate was taken into account variability for the two intervals remained rather stable across the two speaking styles, i.e., the increase in C and V variability (standard deviations) can largely be accounted for by the overall increase in the durations of the intervals.

We conclude this section with a note on individual talker data and intelligibility results. The overall goal of clear speech research is to find acoustic-articulatory changes that talkers consistently produce that may underlie enhanced intelligibility. While we have identified some acoustic-phonetic conversational-to-clear speaking style changes in the temporal domain, the data presented in this paper are limited in terms of providing a direct link between the articulatory changes and an increase in intelligibility. Our database does not contain enough talkers for meaningful correlational analyses between acoustic-phonetic variation and variability in intelligibility. Furthermore, given the nature of our materials (sentences and paragraphs), numerous other conversational-to-clear speech articulatory changes were implemented by our talkers (discussed in Smiljanic and Bradlow, 2005, 2008) and it would be difficult to assess which of these changes (or in which combination) contribute to increased intelligibility. Finally, we can only speculate at this point how the changes in the overall temporal organization, as discussed here, contribute to intelligibility. Nevertheless, we discuss some tendencies observed in our database, which will have to be explored in more detail in future work.

The results revealed a large amount of variability in individual results for various acoustic measurements and their relation to intelligibility. Nevertheless, some acoustic measurements seem correlated with intelligibility. For instance, larger lengthening of C intervals in short sentences and SP, and of V intervals in short sentences corresponds to increased intelligibility. A similar relationship with intelligibility holds for a speaking rate decrease in short sentences and SP, for an increase in the number of C intervals in short sentences and LP, and for V intervals in LP. However, these same measurements do not exhibit a systematic relation to intelligibility in other instances, i.e., with other materials. Larger lengthening of C intervals for LP and of V intervals for SP and LP, for instance, does not result in increased intelligibility. An insufficient number of data points does not allow for a more systematic analysis of these trends. Furthermore, on some acoustic measures, similar amounts of change affect intelligibility differently for various talkers. For instance, an increase of 5 and 11 C intervals in short sentences

results in an intelligibility gain of 35 and 5 percentage points, respectively, while an increase of 24 C intervals results in only 4% point intelligibility increase.

While it is difficult to ascertain the relationship between these acoustic cues and intelligibility due to the multidimensional clear speech changes in these data, nonetheless, it is likely that some of these clear speech adjustments may contribute to enhanced intelligibility. It seems also likely that a combination of different clear speech changes, rather than each one separately, result in increased intelligibility. Furthermore, clear speech probably varies for different materials and communicative demands, i.e., salient and systematic clear speech changes in short sentence production may not be implemented in longer and more complex materials or in spontaneous speech. Clear speech adjustments may also vary across the portions of longer utterances (similar to the intelligibility variability across different portions of the two paragraphs). Finally, even if a direct link between some acoustic cue (e.g., increased F2 for front vowels) and intelligibility is found to be reliable for all talkers and listeners, it still remains to be seen how this cue is used in running speech where linguistic processing and goals differ. Many of these issues remain to be explored in future work.

IV. SUMMARY AND DISCUSSION

This study investigated the effect of hyperarticulated, intelligibility-enhancing clear speech on temporal characteristics in short sentences and paragraph-length utterances. The major goals were to explore whether temporal restructuring at the level of segmental intervals accompanies changes in speaking style across materials of varying complexity. Moreover, we sought to extend previous findings on clear speech intelligibility by testing the maintenance of clear speech across paragraph-length utterances.

The results of sentence-in-noise listening tests showed a consistent clear speech intelligibility gain for short sentences and across the two paragraphs indicating that spontaneously produced clear speech enhances intelligibility for all material types. The data provided strong evidence that talkers successfully maintained clear speech articulatory modifications across longer stretches of speech such that intelligibility was increased in all portions of the paragraphs. Previous clear speech research largely focused on intelligibility of isolated vowels, words, and short sentences (Picheny *et al.*, 1986; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Bradlow and Bent, 2002; Bradlow *et al.*, 2003; Ferguson, 2004; Ferguson and Kewley-Port, 2002). The current results expand on these findings by demonstrating that clear speech modifications can increase the intelligibility of longer and more complex utterances. This finding is an important step toward connecting laboratory-based research on variability in intelligibility to more naturalistic communicative settings where speech typically involves longer utterances with highly variable phrase types.

The present data also provide some insight into the underlying organizational framework for the now well-established decrease in speaking rate for naturally produced clear speech. The decrease in speaking rate was achieved through equal lengthening of consonant and vowel intervals in clear speech: %C and %V remained stable across speaking styles. Furthermore, clear speech slowing down resulted in the increased number of pauses (PPs) and of V and C intervals that were dropped or coarticulated in conversational speech. The results here replicate the findings that the overall speaking rate decrease in clear speech was achieved by a combination of individual segment lengthening and an increase in the number of segments and pauses (Pichney *et al.*, 1986; Krause and Braida, 2004; Ferguson and Kewley-Port, 2002; Smiljanic and Bradlow, 2005, 2008). They also expand the previous findings by looking at the variability of C and V intervals in a more systematic and detailed way across longer utterances. The overall effect of the conversational-to-clear speech style modification, thus, appears to involve some temporal restructuring in how speech unfolds in time as seen in the increased number of C and V intervals (and segments) and of pauses. This restructuring can be seen in the realizations of new articulatory targets that enhance segmental and prosodic compositions of the message. Alternatively, the restructuring could be viewed as a reduction in clear speech of articulatory target undershoot and target overlap present in conversational speech.

The connection between the temporal organization change from conversational to clear speech and their relation to increased intelligibility have yet to be fully established. Krause and Braida (2002), for instance, found that clear speech produced at normal/conversational speaking rates increased intelligibility, suggesting that the temporal changes are not necessary conditions for enhanced intelligibility. However, this does not mean that, when present, the temporal changes of slow clear speech are not important for the intelligibility benefit. Listeners may use the temporal cues when available either to further enhance intelligibility or to process the speech signal in a different manner (weighing available cues differently). The present data suggest several possible temporal-related mechanisms that may underlie the intelligibility increase. First, a decrease in speaking rate may be related to greater articulatory precision as seen, for instance, in an increased distance among the contrastive vowel categories (Smiljanic and Bradlow, 2005). Similarly, slower speaking rate allows for the lengthening and insertion of the short segments that were dropped or coarticulated with surrounding sounds in conversational speech. In addition to increasing the accuracy of the phoneme identification, these two effects conspire to make syllable and word structure more salient, presumably increasing accuracy in lexical access and word recognition.

Second, temporal restructuring through increased phrasing may further contribute to higher clear speech intelligibility compared to conversational speech intelligibility. In the present paper, PPs were crudely quantified as occurring between pauses. The results showed that the number of PPs was increased in clear speech. Moreover, informal observations showed an increase in the number of intermediate (ip) and IPs within PPs as well, i.e., a number of smaller PPs that did not coincide with the pauses were produced as well. These phrases are often marked by intonation patterns (F0), phrase-initial and -final lengthening and reduced coarticula-

tion between phonemes across boundaries (Wightman *et al.*, 1992; Keating *et al.*, 2003; Byrd *et al.*, 2000). There is evidence that the acoustic correlates of these phrases, as well as of word-level stress, aid listeners in lexical segmentation, i.e., in finding word boundaries and in resolving lexical competition (Cho *et al.*, 2007; Cutler and Otake, 1994; Fougeron, 1999; Christophe *et al.*, 2004). Christophe *et al.* (2004) found that monosyllabic words were accessed faster in a two-word sequence when the two words belonged to two separate phonological phrases than when they belonged to the same phrase. Recently, Cho *et al.* (2007) showed that listeners were better at resolving lexical ambiguity in two-word sequences when the initial syllable of the second word originated in a higher PP, i.e., intermediate phrase versus word phrase. The authors argued that under the conditions of lexical competition, domain-initial strengthening (longer and more extensive articulatory gestures) helped listeners derive the correct segmentation. These studies, among other, showed that prosodically driven phonetic detail helps listeners in segmenting speech and resolving lexical ambiguity. Similarly, the results of this study suggest that increased phrasing in addition to the increased segmentability (larger number of C and V intervals) provide contexts in which lexical segmentation and lexical access are improved for clear speech. Such detailed prosodic analyses were, however, beyond the scope of this study and will be pursued in future work.

Finally, an important finding of this study was that, in contrast to the increased duration variability as reflected in larger standard deviations, coefficients of variation for both C and V intervals remained unchanged across clear and conversational speech. Similar coefficient of variation stability for speech rhythm across speaking rates was found by Dellwo *et al.* (2004). This result echoes the findings in Smiljanic and Bradlow, 2008 that the proportional distances between the "short and long" vowel categories and between the voiced and voiceless stop categories remained unchanged across the two speaking styles cross-linguistically. In that study, we argued that duration was not a dimension of vowel and stop voicing category contrast enhancement. Rather, the language-specific pronunciation norms along this dimension were maintained in clear and conversational speech. While accumulated production and perception results from previous studies showed that fine grained phonetic detail of segmental duration encodes information from various linguistic levels and in turn governs speech comprehension (Cho *et al.*, 2007; Cutler and Otake, 1994; Fougeron, 1999; Christophe *et al.*, 2004), these data suggest that the same relative durations for these effects should be maintained across changes in speaking styles and rates. In other words, in order for listeners to successfully interpret the speech signal and derive information about linguistic structure (lexical access, segment identification, prosodic structure, etc.), they may rely on local timing relations, which can be measured as durational ratios. These results support the notion that the mechanism guiding articulatory timing may be inherently relational and that both speech production and perception rely on proportional stability in the temporal domain. Finally, these results suggest that increased intelligibility of clear speech may be attributed to prosodic structure enhancement (increased phrasing and enhanced segmentability) in combination with stable global temporal properties.

ACKNOWLEDGMENTS

We are grateful to Josh Viau for assistance in data collection. We also thank Christine Shadle and three anonymous reviewers for helpful suggestions concerning the research reported here. This research was supported by Grant No. NIH-R01-DC005794 from NIH-NIDCD.

[1]It is important to note that fast-to-slow speaking rate changes, and conversational-to-clear speech modifications may not involve the same articulatory scaling mechanisms and the similarities and differences between the results of the various studies should, therefore, be interpreted with caution.

[2]Some pauses were eliminated due to the segmentation criteria (paragraphs were segmented into the sentence-length utterances in part where pauses occurred) prior to the leveling procedure. However, a few pauses remained within clear speech sentences, which could have affected their average rms amplitude (lower overall for clear than for the conversational sentences). Nevertheless, across all materials, the perceived subjective loudness remained constant (as determined by the author and the experimenters who tested the participants). Any existing difference may have affected the SNR during spoken portions of the sentences (better for clear speech), which could have artificially inflated the clear speech intelligibility benefit. However, this effect would have occurred only for a few sentences and the overall increase in clear speech intelligibility due to this factor is rather small.

Boucher, V. J. (**2002**). "Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates," Percept. Psychophys. **64**, 121–130.

Boersma, P., and Weenink, D. (**2006**). PRAAT: Doing phonetics by computer (Version 4.4.16). Retrieved from http://www.praat.org/ (Last viewd 12/27/2007).

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for learning-impaired children: Sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Bradlow, A. R., and Alexander, J. (**2007**). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**, 2339–2349.

Byrd, D., Kaun, A., Narayanan, S., and Saltzman, E. (**2000**). "Phrasal signatures in articulation," in *Lab. Phonology 5: Acquisition and the Lexicon*, edited by M. B. Broe and J. B. Pierrehumbert (Cambridge University Press, Cambridge), pp. 70–87.

Byrd, D., and Saltzman, E. (**2003**). "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening," J. Phonetics **31**, 149–180.

Byrd, D., Krivokapic, J., and Lee, S. (**2006**). "How far, how long: On the temporal scope of prosodic boundary effects," J. Acoust. Soc. Am. **120**, 1589–1599.

Cho, T. (**2005**). "Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English," J. Acoust. Soc. Am. **117**, 3867–3878.

Cho, T. (**2006**). "Manifestation of prosodic structure in articulation: Evidence from lip movement kinematics in English," in *Laboratory Phonology 8: Varieties of Phonological Competence*, edited by L. Goldstein (Mouton de Gruyter, Berlin, New York), pp. 519–548.

Cho, T., McQueen, J. M., and Cox, E. (**2007**). "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English," J. Phonetics **35**, 210–243.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. (**2004**). "Phonological phrase boundaries constrain lexical access: I. Adult data," J. Mem. Lang. **51**, 523–547.

Cole, J., Kim, H., Choi, H., and Hasegawa-Johnson, M. (**2007**). "Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from radio news speech," J. Phonetics **35.2**, 180–209.

Cutler, A., and Otake, T. (**1994**). "Mora or phoneme? Further evidence for

language-specific listening," J. Mem. Lang. **33**, 824–844.
Dellwo, V., Steiner, I., Aschenberner, B., Dankovičová, J., and Wagner, P. (**2004**). "The BonnTempo-Corpus & BonnTempo-Tools: A database for the study of speech rhythm and rate," in Proceedings of the Eighth ICSLP, Jeju Island, Korea.
de Jong, K. (**1995**). "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," J. Acoust. Soc. Am. **97**, 491–504.
Ferguson, S. H. (**2004**). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**, 2365–2373.
Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.
Ferguson, S. H., and Kewley-Port, D. (**2007**). "Talker differences in clear and conversational speech: Acoustic characteristics of vowels," J. Speech Lang. Hear. Res. **50**, 1241–1255.
Fernald, A. (**2000**). "Speech to infants as hyperspeech: Knowledge-driven processes in early word recognition," Phonetica **57**, 242–254.
Fougeron, C. (**1999**). "Prosodically conditioned articulatory variations: A review," UCLA Working Papers in Phonetics **97**, 1–74.
Fougeron, C., and Keating, P. A. (**1997**). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. **101**, 3728–3739.
Gagne, J.-P., Rochette, A.-J., and Charest, M. (**2002**). "Auditory, visual and audiovisual clear speech," Speech Commun. **37**, 213–230.
Grabe, E., and Low, E. L. (**2002**). "Durational variability in speech and the rhythm class hypothesis," *Lab. Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 515–546.
Helfer, K. S. (**1998**). "Auditory and auditory-visual recognition of clear and conversational speech by older adults," J. Am. Acad. Audiol **8**, 234–242.
Hirata, Y. (**2004**). "Effects of speaking rate on the vowel length distinction in Japanese," J. Phonetics **32**, 565–589.
Hirata, Y., and Whiton, J. (**2005**). "Effects of speaking rate on the single/geminate stop distinction in Japanese," J. Acoust. Soc. Am. **118**, 1647–1660.
Johnson, K., Flemming, E., and Wright, R. (**1993**). "The hyperspace effect: Phonetic targets are hyperarticulated," Language **69**, 505–528.
Junqua, J. C. (**1993**). "The Lombard reflex and its role on human listeners and automatic speech recognizers," J. Acoust. Soc. Am. **93**, 510–524.
Keating, P. A., Cho, T., Fougeron, C., and Hsu, C. (**2003**). "Domain-initial strengthening in four languages," in *Lab. Phonology 6: Phonetic Interpretations*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 145–163.
Kessinger, R. H., and Blumstein, S. E. (**1998**). "Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies," J. Phonetics **26**, 117–128.
Krause, J. C., and Braida, L. D. (**2002**). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**, 2165–2172.
Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**, 362–378.
Kuhl, P. K., Andruski, J. E., Chistovich, L., Chistovich, I., Kozhevnikova, E., Sundberg, U., and Lacerda, F. (**1997**). "Cross language analysis of phonetic units in language addressed to infants," Science **227**, 684–686.
(**2002**). *Handbook of the International Phonetic Association* (Cambridge University Press, Cambridge).
Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).
Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Netherlands).
Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F. G. (**2004**). "Clear speech perception in acoustic and electrical hearing," J. Acoust. Soc. Am. **116**, 2374–2383.
Low, E. L., Grabe, E., and Nolan, F. (**2000**). "Quantitative characterizations of speech rhythm: "Syllable-timing" in Singapore English," Lang Speech **43**, 377–401.
Miller, J. L., Green, K. P., and Reeves, A. (**1986**). "Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast," Phonetica **43**, 106–115.
Miller, J. L., and Dexter, E. R. (**1988**). "Effects of speaking rate and lexical status on phonetic perception," J. Exp. Psychol. Hum. Percept. Perform. **14**, 369–378.
Miller, J. L., and Volaitis, L. E. (**1989**). "Effects of speaking rate on the perceived internal structure of phonetic categories," Percept. Psychophys. **46**, 505–512.
Moon, S.-J., and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.
Nagao, K., and de Jong, K. J. (**2007**). "Perceptual rate normalization in naturally produced rate-varied speech," J. Acoust. Soc. Am. **120**, 2882–2898.
Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.
Perkell, J. S., Zandipour, M., Matthies, M. L., and Lane, H. (**2002**). "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues," J. Acoust. Soc. Am. **112**, 1627–1641.
Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.
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.
Pickett, E. R., Blumstein, S. E., and Burton, M. W. (**1999**). "Effects of speaking rate on the singleton/geminate consonant contrast in Italian," Phonetica **56**, 135–157.
Ramus, F., Nespor, M., and Mehler, J. (**1999**). "Correlates of linguistic rhythm in the speech signal," Cognition **72**, 265–292.
Schum, D. J. (**1996**). "Intelligibility of clear and conversational speech of young and elderly talkers," J. Am. Acad. Audiol **7**, 212–218.
Sedaris, D. (**2001**). *Me Talk Pretty One Day*, New Ed ed. (Abacus, New York, New York).
Silverman, K. E. A., Beckman, M., Pitrelli, J. F., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (**1992**). "ToBI: A standard for labeling English Prosody," in Proceedings of the 1992 International Conference on Spoken Language Processing, Banff, Canada, pp. 867–870.
Skowronski, M. D., and Harris, J. G. (**2005**). "Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments," Speech Commun. **48**, 549–558.
Smiljanic, R., and Bradlow, A. R. (**2005**). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. **118**, 1677–1688.
Smiljanic, R., and Bradlow, A. R. (**2008**). "Stability of temporal contrasts in conversational and clear speech," J. Phonetics **36.1**, 91–113. http://www.sciencedirect.com/science/journal/00954470. Last viewed 1/03/08.
Smiljanic, R., and Bradlow, A. R. (**2007**). "Clear speech intelligibility: Listener and talker effects," Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrucken, Germany.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
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.
Uchanski, R. M. (**1988**). "Spectral and temporal contributions to speech clarity for hearing impaired listeners," Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.
Volaitis, L. E., and Miller, J. L. (**1992**). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," J. Acoust. Soc. Am. **92**, 723–735.
Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (**1992**). "Segmental durations in the vicinity of prosodic phrase boundaries," J. Acoust. Soc. Am. **91**, 1707–1717.

A linear model of acoustic-to-facial mapping: Model parameters, data set size, and generalization across speakers

Matthew S. Craig[a)]
Institute of Biomaterials and Biomedical Engineering and Oral Dynamics Laboratory, University of Toronto, 160-500 University Avenue, Toronto, Ontario M5G 1V7, Canada

Pascal van Lieshout
Department of Speech Language Pathology, Oral Dynamics Laboratory, Institute of Biomaterials and Biomedical Engineering, and Department of Psychology, University of Toronto and Toronto Rehabilitation Institute, Toronto, Canada

Willy Wong
Institute of Biomaterials and Biomedical Engineering and Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada

(Received 11 September 2007; revised 13 August 2008; accepted 14 August 2008)

The relationship between acoustic and visual speech is important for understanding speech perception, but it also forms the basis behind a type of facial animator, which can predict facial motion during speech given an acoustic input. This relationship was examined by revisiting a linear transformation model of audio-visual speech production. A mathematical model is constructed whereby the visual aspect of speech is reproduced from the acoustic signal via a linear transformation. Unlike previous studies in this area, this paper will address specific aspects of the model as related to the effects of window size for acoustic framing and the critical size of the training set. On average, facial motion is predicted with a correlation of 0.70 to the recorded motion, when the model is trained and then tested on the same subject. This is comparable to previous studies using either similar or different model approaches. Using a model trained on other subjects and then applying it to a new subject resulted in a prediction correlation of 0.65. Furthermore, acoustic windows of 100 ms and a data set of approximately 40 sentences are required for maximum predictability. The results are interpreted in terms of the underlying assumptions of the model.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2982369]

PACS number(s): 43.70.Jt, 43.70.Bk, 43.72.Ct, 43.70.Gr [AL] Pages: 3183–3190

I. INTRODUCTION

Speech is both an acoustic and visual means of communication. People with extensive hearing loss can understand speech to some degree by lip reading (Dodd, 1987), while even normal hearing individuals lip-read constantly while communicating (Summerfield, 1992). The perceptual congruency between acoustics and motion can lead to "talking heads:" acoustic-to-visual models that predict and animate facial motion from acoustic input.

If synthetic animations of a face can be derived from acoustics alone, this would allow the recreation of the talking face with a significantly lower bit-rate when compared to transmitting video images. These talking heads can further be useful in the development of assistive devices for people with communication disorders, like the hearing impaired or for patients with limitations in facial expression, for example as found in those suffering from Parkinson's disease. The rehabilitation of facial movement during speech, along with correcting problems with articulation for voicing, is essential to improve the quality of life of such patients. Using animation as a biofeedback device, patients might regain some functional facial motion during speech.

Current approaches involve the use of hidden Markov models (HMMs) or time delay neural networks (TDNNs) to predict motion from acoustics, as they are well-suited to model the nonlinear and time-asynchronous relationship between acoustic and visual speech. These models are designed to capture the effects of coarticulation, which would imply that the position of the face during speech cannot be predicted solely from the immediate acoustic parameters, but will also depend on values which precede and follow the current time instant. Accordingly, models will use as much as nine frames of acoustic information (i.e., the acoustic information in a time-window of nine frames of video information) to predict the facial position at a single frame (Fu *et al.*, 2005; Savran *et al.*, 2006). In the case of HMMs, which rely on artificial segmentation of the data, acoustic parameters are converted into phonemic units (i.e., a specific sound pattern for a given language) and the visual data into visemes (i.e., the basic facial configurations of visual speech). Animations are constructed by mapping each phoneme to its corresponding viseme. The motion between adjacent visemes is then interpolated for a smooth animation.

TDNNs search for a nonlinear transformation function that maps acoustics directly to facial motion, but also use acoustic information from adjacent frames to model coarticulation. The result is dependent upon the topology of the net-

[a)]Electronic mail: p.vanlieshout@utoronto.ca

work, the number of layers and nodes, and the initial estimates of node weights. Consequently, these models can vary widely between similar studies from as few as ten hidden units (Yehia *et al.*, 2001) to as many as 600 (Massaro *et al.*, 1999). Values between 25 and 100 are commonly used (Agelfors *et al.*, 1999; Hong *et al.*, 2002; Zelezny *et al.*, 2006). Generally, phonemically richer speech data sets require more nodes, thereby exacerbating the "black-box" nature of TDNNs, and requiring considerable computational resources and time in the training phase (Hong *et al.*, 2002).

Another popular method uses an audio-visual codebook (AVC) (Kakumanu, 2002; Savran *et al.*, 2006). The salient acoustic and visual features of each audio-visual frame are extracted and stored as a pair in a codebook. Input acoustic features are compared to the acoustic features stored in the codebook until a "nearest match" is found. The visual partner of the nearest match is then used as the predicted facial motion for that frame. AVCs can also model coarticulation by using adjacent acoustic frames to identify the best nearest match (Kakumanu, 2002).

A much simpler model of audio-visual speech production involves the use of a multilinear regression (MLR) model (Yehia *et al.*, 1998; Barker and Berthommier 1999; Jiang *et al.*, 2000a, 2000b). MLRs relate acoustics to motion through a simple linear matrix operation. They perform frame-to-frame transformations from acoustic data to facial position. As a result, coarticulation cannot be explicitly modeled. MLRs do not solve for the transformation iteratively as do neural nets, nor do they require initial estimates. Although approximate, MLRs are thus substantially simpler than TDNNs, HMMs, and AVCs in terms of computational expense and objective physical interpretation.

Previous work has indicated that approximately 70% of facial motion can be recovered using a linear regression model, although these studies have been limited to small or restricted speech databases such as monosyllabic words (Jiang *et al.*, 2002), bisyllabic words (Berthommier, 2003) or a small number of sentences (Yehia *et al.*, 2001). Furthermore, these studies are based on a small number of subjects. The purpose of this study is to revisit the use of a linear regression model as a foundation for animation of human speech. Specifically, in extending previous work in this area, we use a significantly larger subject pool and a large set of phonetically varied stimuli to test the effect of the following model aspects on facial motion predictability:

(1) Acoustic window size: typically about 10–16 ms of acoustic data is used to predict facial position at a given time. However, averaging a larger section of acoustic data may increase predictability by capturing adjacent context and smoothing facial motion trajectories.
(2) Size of the data set: small training databases will create acoustic-to-motion transformations that are too narrowly defined and may not generalize well to new sentences. Our goal is to determine a minimum size of dataset for satisfying predictability.
(3) Speaker independence: to see whether we can predict speech motions of subjects who are untrained in the system.

By examining these variables with a significantly larger subject pool, our goal is to present MLR as a conceptually and computationally simpler, yet comparably effective, basis for speech animation as compared to HMMs, TDNNs, and AVCs.

II. METHOD

A. Subjects

Sixteen subjects, eight adult males (M1–M8) and eight adult females (F1–F8), participated in the study. All subjects were over the age of 18 and spoke English as a first language, with the exception of subject M8, whose first language was Tamil but spoke English fluently. None of the subjects had a reported history of speech, language, or hearing problems and all gave informed consent before participating in the study.

B. Stimuli

Ninety sentences were used, 45 from the TIMIT sentences database and 45 from the HARVARD sentences database. The TIMIT sentences provide a rich mixture of phonemes (Brugnara, 2001) while the HARVARD sentences consist of phoneme proportions that are representative of their occurrence in the English language (Logan *et al.*, 1989). Once intersentence pauses were removed from the data, the 90 sentences on average (depending on individual speech rate differences) contained 3 min and 20 s of continuous speech. To test how the size of a database would affect predictability, we ran predictions with sets containing different numbers of sentences, varying from a minimum of 2 to a maximum of 60 sentences (see Sec. III).

C. Audio processing

Speech was recorded using a WMS-PRO VHF condenser microphone with a sampling rate of 32 kHz. Audio time series were split into frames of 100 ms, with a 63.3 ms overlap. A shorter frame length of 10–50 ms is common in many speech coding systems (Kabal, 2003), including facial animators (Jiang *et al.*, 2000a, 2000b, Yehia *et al.*, 1998; Massaro *et al.*, 1999). We found that facial motion predictability was improved with a longer frame in our time-invariant model. Frame sizes ranging from 20 to 150 ms were tested for the model described here (see Sec. III).

Each frame was multiplied by a Hamming window to reduce spectral distortions and the acoustic information was represented by an array of 65 parameters: The root mean squared energy, 16th order linear predictor coefficients (LPC), 16th order line spectral pairs (LSP), and the first derivatives of the previous two sets. LSPs are derived from LPCs (Jiang *et al.*, 2000a) and are closely related to the formant values and bandwidths during speech (Fig. 1).

LPC and LSP coding is common in facial animators (Jiang *et al.*, 2000a, 2000b; Yehia *et al.*, 1998; Massaro *et al.*, 1999). The use of time derivatives can be justified because they reflect the behavior of formant transitions which affect the perception of speech. Thus, the time derivative of LPC and LSP may also have a systematic effect on

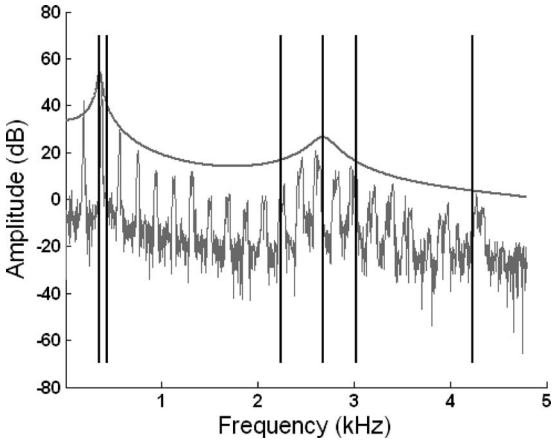


FIG. 1. The power spectrum of an acoustic frame during the production of /i/ by a male speaker, with a 16th order LPC filter response (envelope) and first six LSPs (dark vertical lines).

characteristics of facial motion. Zelezny *et al.* (2006) used 52 parameters, including the first, second, and third derivatives of mel frequency cepstral coefficients, in their audio-visual speech synthesis. Time derivatives have also proven to be useful in speech recognition systems (Chou and Juang, 2003).

D. Video processing

Speech motions were recorded in three dimensions using two JVC DVL9800U digital video cameras placed approximately 1 m in front of the subjects' faces, to the left and right. The cameras have a native shutter speed of 30 frames per second and a digital resolution of 720×480 pixels. Shutter speed was increased to 60 frames per second using interlaced scan mode, which effectively halves the vertical resolution.

Two 250 W ultraviolet black lights, fixed to the ceiling, were used to illuminate 13 glow-in the-dark stickers, which are placed on the subjects' faces at anatomical points of interest (see Fig. 2, left). Each sticker is 3.0+/-0.5 mm in diameter.

Four gestures (see Fig. 2, right) defined as the distance between two markers, were selected that represent facial motion in relation to relevant speech features: jaw, lip corners (LC), upper lip (UL), and upper lip / lower lip distance (UL/LL). Jaw and UL were measured in relation to a fixed point on the subjects' foreheads. The UL gesture involves motion

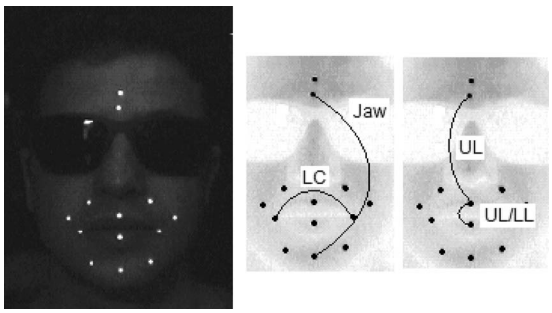


FIG. 2. A subject with illuminated markers (left). The four gestures used in this study (right).

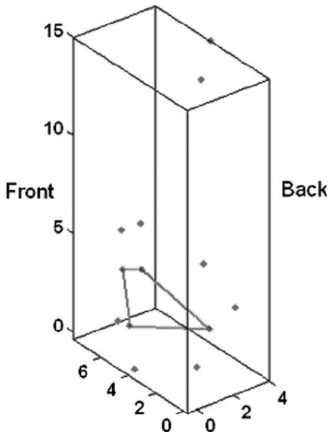


FIG. 3. Extracted 3D marker positions showing the oral area between the lips. The polygon connects the four points on the lips.

in the depth and vertical dimensions for protrusions and bilabial closures, the LC gesture covers mainly motion in the horizontal dimension, while UL/LL and jaw involve mainly vertical motions. Similar facial measures have been reported in the literature for linear regression studies as well as other visual speech synthesis models (Barker and Berthommier, 1999; Curinga *et al.*, 1996; Jiang *et al.*, 2000a, 2000b 2002; Kakumanu *et al.*, 2006). The gestures were selected for clarity and simplicity, and represent anatomical motions that may be of interest during speech.

The recorded video from both cameras was processed using a specialized software suite, APAS, developed by Ariel Dynamics, Inc. (Trabuco Canyon, USA). APAS converts the locations of each marker in each video frame into a three-dimensional location, creating a time-series that can be read into Matlab®. This process is known as digitization. Figure 3 depicts the digitized three dimensional (3D) positioning of markers. For clarity, a polygon is added connecting the four points on the lips.

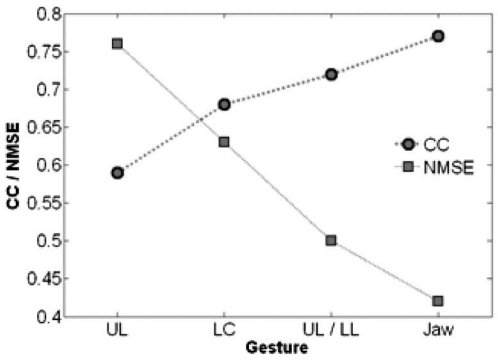
After digitization, gesture trajectories were smoothed using a Butterworth low-pass filter with a cut-off frequency of 16 Hz. Virtually all speech motions fall under 8 Hz (Craig *et al.*, 2007).

E. System resolution

A detailed description of the system's resolution and accuracy can be found elsewhere (Craig *et al.*, 2007). The video cameras and digitization software are capable of resolving motions of amplitude as low as 0.7 mm with excellent accuracy. To be conservative, kinematic data from individual sentences with amplitudes less than 1 mm were discarded from this study. This was only the case for some instances of the UL gesture, which has an average amplitude of 1.2 mm for adult subjects during normal speech (Craig *et al.*, 2007).

F. Multilinear Regression

A MLR approach is used to map acoustic speech to facial position. The MLR defines a transformation, **T**, that multiplies each acoustic frame, **A**, to predict each of the four



Gesture	CC (STD)	NMSE (STD)
Jaw	*0.77 (0.06)*	*0.42 (0.07)*
UL/LL	*0.72 (0.07)*	*0.50 (0.04)*
LC	*0.68 (0.10)*	*0.63 (0.12)*
UL	*0.59 (0.11)*	*0.76 (0.13)*
Mean	***0.69 (0.06)***	***0.58 (.07)***

FIG. 4. Overall gesture CCs and NMSEs across sentences.

gestures of motion of each visual frame, **V**. There is one transformation function for each gesture. Specifically, for a data set of n frames

$$\mathbf{T} \cdot \mathbf{A} = \mathbf{V}, \tag{1}$$

where

$$\mathbf{A} = \begin{pmatrix} A_{1,1}, \ldots, A_{1,n} \\ A_{2,1}, \ldots, A_{2,n} \\ \ldots \\ A_{p,1}, \ldots, A_{p,n} \end{pmatrix},$$

$$\mathbf{V} = \begin{pmatrix} V_{1,1}, \ldots, V_{1,n} \\ V_{2,1}, \ldots, V_{2,n} \\ \ldots \\ V_{m,1}, \ldots, V_{m,n} \end{pmatrix}. \tag{2}$$

$\mathbf{A_{p,n}}$ represents acoustic parameter p in frame n and $\mathbf{V_{m,n}}$ represents visual gesture m in frame n. "p" ranges from 1 to 65 for the 65 acoustic parameters, "m" from 1 to 4 for the four visual gestures, and "n" is the total number frames in the training set which is approximately equal to 12 000 frames (3 min and 20 sec at 60 frames per second).

In multilinear regression, the optimal transformation, **T**, is obtained by minimizing the squared error between the predicted motion, $\mathbf{P}=\mathbf{T}\cdot\mathbf{A}$ and recorded motion, **V** for all frames.

$$\mathbf{SE} = \sum_{k=1 \text{ to } n} (\mathbf{P}_k - \mathbf{V}_k)^2. \tag{3}$$

T is found by setting the derivative of Eq. (3) to zero. The resulting solution is unique (Ferguson, 1984)

$$\mathbf{T} = [(A) \cdot (A)^T]^{-1}(A) \cdot (V)^T. \tag{4}$$

G. Model evaluation

For a given data set of acoustic and visual data for **S** spoken sentences, the data is separated into training and testing sets. **S**-1 sentences comprise the training set [i.e., are used to define the transformation **T** from Eq. (4)]. The final sentence is used as the test set: **T** is multiplied by the acoustic parameters of each frame of the test repetition, yielding four predicted facial trajectories (time series of distances), one for each gesture. The predicted trajectories are compared to the recorded trajectories for that sentence and gesture by means of a Pearson correlation coefficient (CC) and normalized mean square error (NMSE). This procedure is repeated for all different combinations of training and testing data. CCs relate how well the signal's trajectories match, with a value of 1 being a perfect directional match, while NMSEs relate how well the signal's amplitudes match. We expect a negative correlation between CCs and NMSEs (i.e., a well predicted signal will have a high CC and low NMSE). However, for some predictions it may be the case that directionality is well predicted but scaling is inaccurate, leading to both a high CC and a high NMSE. Unlike CC there are no upper boundaries for NMSE. A zero value obviously means that there are no amplitude differences between the predicted and actual time series.

III. RESULTS

When training and testing was performed for the same subject's data, the overall CC was 0.69 and overall NMSE was 0.58 across all gestures and subjects (Fig. 4).

In most subjects, jaw motion was predicted best across sentences CC=0.77 (0.06), NMSE=0.42 (0.07), while UL (upper lip) motion was predicted worst, CC=0.59 (0.11), NMSE=0.76 (0.13). A one-way ANOVA revealed that the differences in CC between all gestures was significant [$F(3,15)=40.5$, $p<0.001$]. The distribution of CCs, averaged across gestures, is shown in Fig. 5. A sample of recorded and

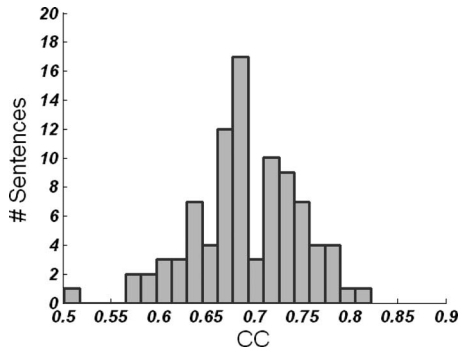


FIG. 5. Distribution of sentence mean CCs, averaged across subjects and gestures.

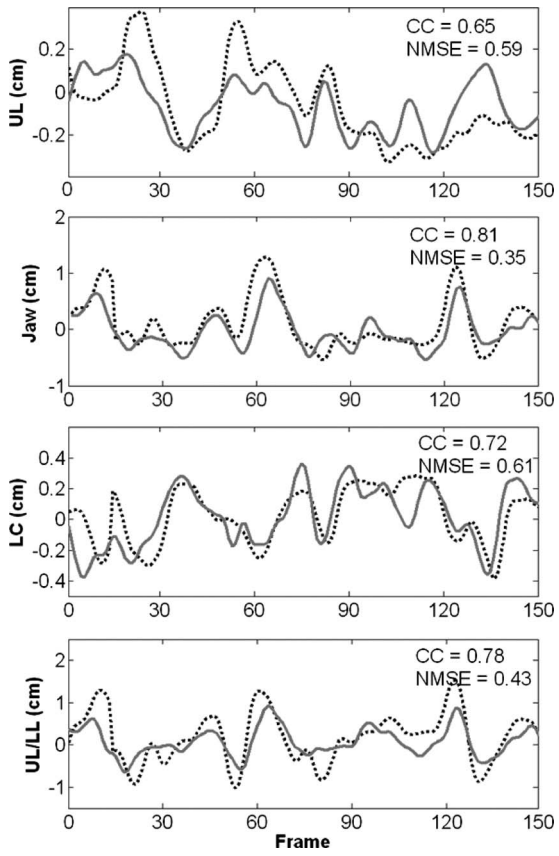


FIG. 6. Predicted (solid line) vs actual (dotted line) trajectories for the four different gestures (jaw, UL, LC, and UL/LL) for the sentence "The museum hires musicians every evening" from subject F3. These data illustrate different combinations for CC and NMSE values.

predicted trajectories for the four gestures is shown in Fig. 6, illustrating different combinations of CC and NMSE values for a single sentence and subject.

Scatter plots in general revealed high negative correlations between CC and NMSE, indicating that in most cases good predictions yielded relatively high CCs and correspondingly low NMSEs. Figure 7 shows a typical example for subject M6. The mean correlations between CC and NMSEs across subjects are reported in Table I. It was seldom the case that both CC and NMSE were high indicating that there is no systematic scaling error for the predicted signal. This implies that either CC or NMSE alone is sufficient to report

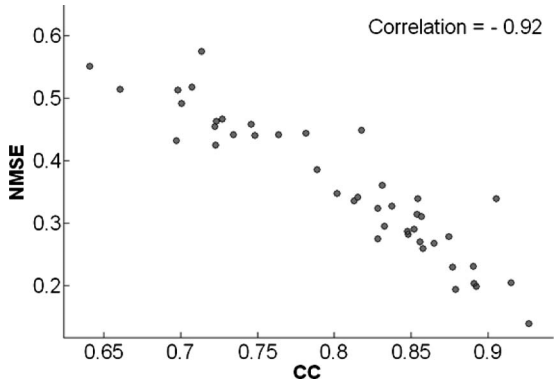


FIG. 7. Correlation between CC and NMSE for jaw predictions for subject M6.

TABLE I. Correlation between CCs and NMSEs across subjects.

UL	LC	UL/LL	Jaw
−0.62 (0.21)	−0.76 (0.12)	−0.92 (0.05)	−0.91 (0.05)

our results. For the sake of comparison with other studies, we will present our results mainly as CCs. Only in the case of UL predictions, and rarely for LC were there significant outliers to this trend for a few specific sentences. For example, for the sentence "Nothing is as offensive as innocence," in which the UL has no functional use, predicted trajectories were scaled out of proportion often leading to extremely high NMSEs and moderate CCs, as shown in Fig. 8.

Figure 9 (top) depicts the sensitivity of predictability to acoustic frame size. Predictability was highest at a frame size of 100 ms, justifying the use of the 100 ms frame size for all subjects in these analyses. Figure 9 (bottom) shows the dependence of predictability on the size of the training database. At a training dataset of 37 sentences, predictability has reached 99% of its maximum value, but as shown in this figure, differences in predictability for data sets larger than 20 are very small.

Another purpose of the present study was to determine whether our system would be successful in predicting facial motion of a subject whose data were not used for training the transformation matrix. This can be tested in two ways within each gender: the first involves using the data from all subjects of one gender, excluding the test subject, to define a single transformation, which is then used on the test data. This method is called "multisubject transformation." The second possibility is to create individual transformations for subjects and then average them together into one transformation, which is applied to the remaining subject's data. This method is called "subject averaged transformation."

For the multisubject transformation, 30^1 sentences from seven subjects were combined into a 210-sentence training set. This training set was used to define a transformation, which was then tested on the 30 sentences of the remaining test subject. This process was repeated eight times for all combinations of training/testing subjects per gender.

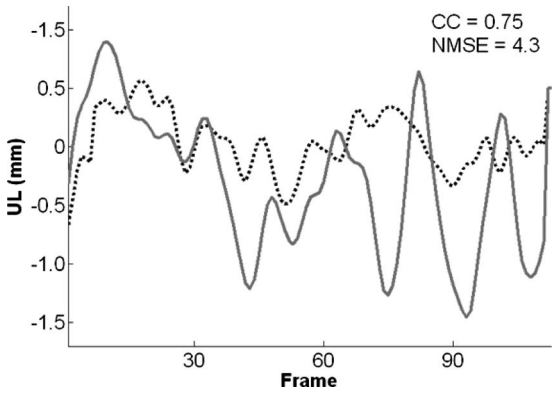


FIG. 8. Overscaled predictions for UL from subject M5 for the sentence "Nothing is as offensive as innocence." As in Fig. 6, solid lines depict predicted trajectories and dotted line depicts actual trajectories.

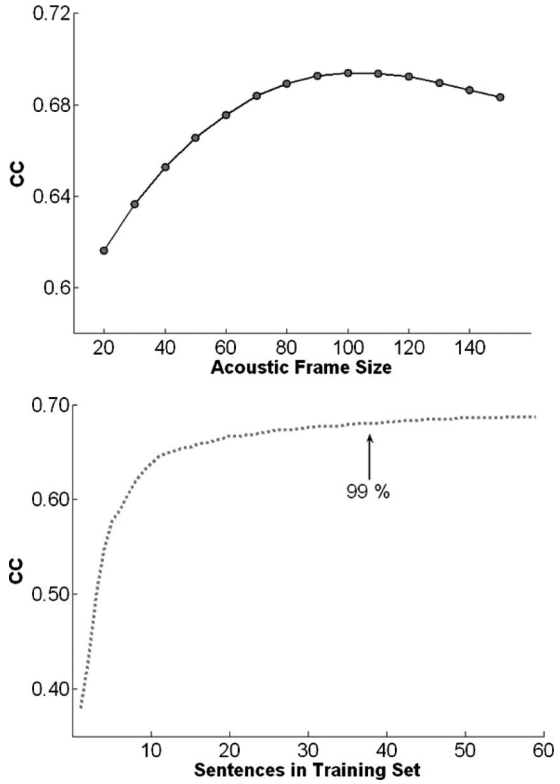


FIG. 9. Top: Sensitivity of CCs to acoustic frame size, using 89-sentence training sets, averaged across subjects. Bottom: Sensitivity of CC to the number of sentences in the training data set using 100 ms acoustic frame, averaged across subjects.

For the subject averaged transformation, the same 30 sentences from each subject were used to define transformations for each subject. Seven transformations were then linearly averaged into one transformation, which was tested on the 30 sentences of the remaining test subject. This process was repeated eight times for all combinations for training/testing subjects per gender.

Results for the multisubject transformation and subject averaged transformation are given in Table II. Results using the same subject for testing and training are also given under "single subject." Figure 10 shows the individual gesture differences.

The subject averaged model generalizes well across subjects of the same gender, with average CCs dropping from 0.71 to 0.66 for males and from 0.68 to 0.60 for females. The multisubject model did not perform as well, with drops in CC from 0.71 to 0.59 for males and 0.68 to 0.52 for females.

Differences in CC were largest for UL, particularly for female subjects, indicating that the UL gesture may be the least generalizable and this is probably also a reflection of the relatively poor UL prediction in females in general.

TABLE II. A comparison of subject-dependent and subject-independent results.

	Male	Female
Single subject	0.71	0.68
Multisubject	0.59	0.52
Subject averaged	0.66	0.60

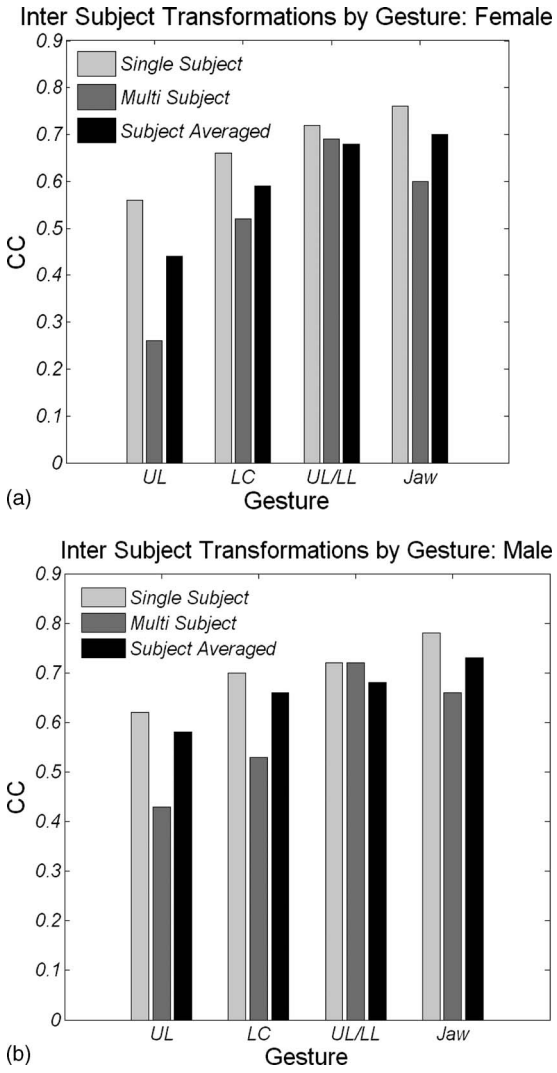


FIG. 10. Single subject, multisubject, and subject averaged transformations, separated by gesture.

IV. DISCUSSION

Our results are consistent with the previously obtained results from smaller scale studies using regression modeling, that approximately 70% of facial motion is recoverable (Yehia *et al.*, 1998; Barker and Berthommier 1999; Jiang *et al.*, 2000a, 2000b). We have confirmed these results for a much larger subject pool and highly varied speech database.

The findings further show that acoustic windows of 100 ms and data sets of approximately 40 sentences are required to achieve maximum predictability. In addition, we showed that by using subject averaged transformations, the facial motions of new subjects (i.e., not included in the training phase) can be predicted with only a small drop in accuracy.

An important limitation with the MLR approach is that it assumes that a linear, time-invariant mapping exists between acoustic information and facial position. This is demonstrably untrue even for most simplistic speech. Schroeder (1967), modeling the vocal tract as a tube, showed that a unique area function could not be determined by the formants of the output acoustics. Computer simulations have shown that different vocal tract shapes can lead to almost

identical values of the first three formants (Atal *et al.*, 1978). A recent paper describing a new mathematical approach called MIMICRI deals with this issue by applying continuity constraints in order to recover articulator motion from acoustics, "even if the mapping from articulatory positions to acoustics is not known, and possibly, even if the mapping is many-to-one" (Hogden *et al.*, 2007, p. 379). This approach may hold some promise for future applications in the area of speech recognition and animation.

The situation for our model becomes more complicated for vocal-tract articulator position rather than facial articulator position because intra-oral articulator motion (e.g., tongue) is hardly reflected in the face (except when it directly correlates with jaw motion; see below). The articulations of /t/, /d/, and /n/ (as in /toe/, /doe/, and /know/) can look almost identical in the face but sound very different. Facial articulator position for a given sound can vary widely for a single phoneme across contexts. For instance, the tongue and jaw can act in a coordinated manner to achieve a single acoustic goal in a number of different ways, referred to as "coarticulation," or "functionally-organized goal-oriented behavior" (Fletcher and Harrington, 1999). Moreover, the individual contribution of articulators to produce a phoneme can vary between repetitions of the same utterance, and may vary considerably between sessions for the same subject and context (Alfonso and Van Lieshout, 1997).

A number of methods have been used to capture the time-varying nature of articulation in facial animators using other models such as AVCs (Kakumanu, 2002), TDNNs (Massaro *et al.*, 1999; Savran *et al.*, 2006), and HMMs (Yamamoto *et al.*, 1998; Xie and Liu, 2007). Despite the limitations of MLR, our results are comparable to other systems using non-linear, asynchronous models. Xie and Liu (2007) compared six different HMMs for audio-visual speech synthesis, the most effective model producing average correlations of 0.696. Zelezny *et al.* (2006) using a 39-input ANN achieved a highest correlation of 0.622 when predicting LC, UL/LL and UL. Fu *et al.* (2005) were able to achieve a higher average correlation of 0.83 for jaw, UL/LL and LC with a HMM that used five frames of video information (captured at 60 fps) to account for backwards and forwards coarticulation. Savran *et al.* (2006) using a time delay neural network with nine frames of context achieved correlations of 0.72. They were then able to further increase predictability by using text information as an additional input.

In our system, the use of long acoustic frames acted to smooth over local context and coarticulation effects, thereby improving predictability. Particularly interesting is the fact that the subject averaged transformations generalized well to new subjects (CC=0.66 for males, 0.60 for females) even though the number of subjects used to obtain the average transformation was relatively small (7), and predictability will likely improve with a larger pool. Savran *et al.* (2006) attempted similar subject-independent predictions by concatenating the audio-visual speech data of 14 subjects into an audio-visual codebook, and used a look-up algorithm to predict facial motion of a novel subject. However, the average correlation of 0.72 using a training database of 125 sentences was obtained using text information as an additional input.

The poorer results of the multisubject transformations are most likely due to the fact that all intersubject data, including large differences in orofacial dimensions and absolute values of formant frequencies, are being concatenated into the training set before the minimum squared error algorithm is performed. With the subject-averaged transformations, intersubject differences are being concatenated by an averaging process after each minimum squared error transformation is derived. This second process is less sensitive to variations in the total data set and will lead to better general predictions.[2] Savran *et al.* (2006) avoided this issue by associating acoustic information from multiple speakers to visual information from a single speaker in a codebook algorithm, using a specific weighting system. In their system, the acoustic information of a given speaker is independently compared to that of 14 other speakers before choosing the appropriate viseme.

Our subject averaged model reduces these differences by first creating individual transformations and then averaging them. This suggests that the important structural similarities across subjects for the mappings are related to the general acoustic-to-articulatory trends, e.g., that F1 is inversely related to tongue height while F2 is directly related to tongue advancement (Kent and Read, 1992). Both height and advancement of the tongue are partially controlled by the jaw and will thus be mirrored, to some extent, in the face (Hertrich and Ackermann, 2000; Stone and Vatikiotis-Bateson, 1995).

The speaker-independence of MLR modeling may make it useful for animation in the context of rehabilitation purposes. Berthommier (2003) has shown that animations based on linear regressions are realistic enough to improve speech perception, although that study was limited to one-syllable or two-syllable words with overt syllabic structures. We generated some preliminary crude point-animations and have found them to look surprisingly realistic despite mediocre correlation coefficients. This provides hope to its use in future multimedia/rehabilitative applications, although more objective testing with realistic animations is necessary.

V. CONCLUSIONS

We have shown that facial motion can be predicted from acoustics using a linear, time-invariant model with accuracies comparable to that of more computationally intensive models. Our results were generated by 16 subjects speaking 90 sentences each.

While HMMs, TDNNs, and AVCs use adjacent acoustic frames to model coarticulation, this is not possible using the technique of MLR. We compensated for this by using an enlarged acoustic window of 100 ms, which then improved predictability. To train our system effectively, approximately 40 sentences (less than 2 min of continuous speech) are required. Further, our model is capable of predicting the speech motions of subjects untrained in the system with only a small drop in predictability. This work demonstrates that MLR may be used as a general framework to create facial animations for multimedia or rehabilitative purposes.

ACKNOWLEDGMENTS

This research was supported by the Natural Sciences and Engineering Research Council (NSERC) and undertaken, in part, thanks to funding from the Canada Research Chairs Program. The authors wish to thank Aravind Namasivayam and Mark Noseworthy for their support and help in preparing the manuscript.

[1]As mentioned, predictability differences for datasets larger than 20 were very small and close to the 99% predictability level found for 37 sentences and 30 sentences were chosen as a practical number (all sentences were recorded in groups of 15) for the purpose of this evaluation.

[2]We thank an anonymous reviewer for highlighting this aspect.

Agelfors, E., Beskow, J., Granstrom, B., Lundenber, M., Salvi, G., Spens, K., and Ohman, R. (**1999**). "Synthetic visual speech driven from auditory speech," Proceedings of Audio-Visual Speech Processing, Santa-Cruz, pp. 123–127.

Alfonso, P. J., and Van Lieshout, P. (**1997**). "Spatial and temporal variability in gestural specification," in *Speech Production: Motor Control, Brain Research and Fluency Disorders*, edited by W. Hulstijn, F. Peters, and P. van Lieshout (Elsevier Science, Amsterdam), pp. 151–160.

Atal, B. S., Chang, J. J., Mathews, M. V., and Tukey, J. W. (**1978**). "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting technique," J. Acoust. Soc. Am. **63**, 1535–1555.

Barker, J. P., and Berthommier, F. (**1999**). "Estimation of speech acoustics from visual speech features: A comparison of linear and non-linear models," Proceedings of Audio-Visual Speech Processing, Santa-Cruz, pp. 112–117.

Berthommier, F. (**2003**). "Direct synthesis of video from speech sounds for new telecommunication applications," in Smart Object Conference, Grenoble, France.

Brugnara, F. (**2001**). "Model agglomeration for context-dependent acoustic modeling," Proceedings of Eurospeech, Aalborg, pp. 1641–1644.

Chou, B. H., and Juang, W. (**2003**). *Pattern Recognition in Speech and Language Processing* (CRC Press, Boca Raton).

Craig, M., Van Lieshout, P., and Wong, W. (**2007**). "Suitability of a UV-based video recording system for the analysis of small facial motions during speech," Speech Commun. **49**, 679–686.

Curinga, S., Lavaghetto, F., and Vignoli, F. (**1996**). "Lip movements synthesis using time delay neural networks," Proceedings of European Signal Processing Conference, Triest, pp. 999–1002.

Dodd, B. (**1987**). *Hearing by Eye: The Psychology of Lip Reading* (Lawrence Erlbaum Associates, Hillsdale, NJ).

Ferguson, G. A. (**1984**). *Statistical Analysis, Synthesis, and Perception* (Springer, New York).

Fletcher, J., and Harrington, J. (**1999**). "Lip and jaw coarticulation," in *Coarticulation: Theory, Data and Techniques*, edited by W. Hardcastle and N. Hewlett (Cambridge University Press, Cambridge), pp. 164–178.

Fu, S., Gutierrez-Osuna, R., Esposito, A., Kakumanu, P., and Garcia, O. (**2005**). "Audio/visual mapping with cross-modal hidden Markov models," IEEE Trans. Multimedia **7**, 243–252.

Hertrick, I., and Ackermann, H. (**2000**). "Lip-jaw and tongue-jaw coordination during rate-controlled syllable repetitions," J. Acoust. Soc. Am. **107**, 2236–2246.

Hogden, J., Rubin, P., McDermott, E., Katagiri, S., and Goldstein, L. (**2007**). "Inverting mappings from smooth paths through Rn to paths through Rm: A technique applied to recovering articulation from acoustics," Speech Commun. **49**, 361–383.

Hong, P., Wen, Z., and Huang, T. S. (**2002**). "Real-time speech-driven face animation with expressions using neural networks," IEEE Trans. Neural Netw. **13**, 916–927.

Jiang, J., Alwan, A., Bernstein, L., Keating, P., and Auer, E. (**2000a**). "On the correlation between facial movements, tongue movements and speech acoustics," Proceedings of the International Conference on Spoken Language Processing, Beijing, Vol. **1**, pp. 42–45.

Jiang, T., Li, Y., and Chen, H. (**2000b**). "A 1.44 Kbps vocoder based on LSP," Proceedings of the 5th International Conference on Signal Processing, Beijing, pp. 697–701.

Jiang, J., Alwan, A., Bernstein, L., Alwan, A., and Keating, P. (**2002**). "Predicting face movements from speech acoustics using spectral dynamics," Proceedings of the International Conference on Multimedia and Expo, Lausanne, pp. 181–184.

Kabal, P. (**2003**). "Time windows for linear prediction of speech," Version 2. Technical report, Department Electrical and Computer Engineering, McGill University, Montreal.

Kakumanu, P. (**2002**). "Audio-video processing for speech driven facial animation," M.Sc. thesis, Computer Science Department, Wright State University, Dayton, OH.

Kakumanu, P., Esposito, A., Garcia, O., and Gutierrez-Osuna, R. (**2006**). "A comparison of acoustic coding models for speech-driven facial animation," Speech Commun. **48**, 598–615.

Kent, R. D., and Read, C. (**1992**). *The Acoustic Analysis of Speech* (Singular Publishing Group, Inc., San Diego).

Logan, J. S., Greene, B. G., and Pisoni, D. B. (**1989**). "Segmental intelligibility of synthetic speech produced by rule," J. Acoust. Soc. Am. **86**, 566–581.

Massaro, D. W., Beskow, J., Cohen, M. M., Fry, C. L., and Rodriquez, T. (**1999**). "Picture my voice: Audio to visual speech synthesis using artificial neural networks," Proceedings of Audio-Visual Speech Processing, Santa-Cruz, pp. 133–138.

Savran, A., Arslan, L., and Akarun, L. (**2006**). "Speaker-independent 3D face synthesis driven by speech and text," Signal Process. **86**, 2932–2951.

Schroeder, M. R. (**1967**). "Determination of the geometry of the human vocal tract by acoustic measurements," J. Acoust. Soc. Am. **41**, 1002–1010.

Stone, M., and Vatikiotis-Bateson, E. (**1995**). "Trade-oils in tongue, jaw and palate contributions to speech production," J. Phonetics **23**, 81–100.

Summerfield, Q. (**1992**). "Lipreading and audio-visual speech perception," Philos. Trans. R. Soc. London, Ser. B **335**(1273), 71–78.

Xie, L., and Liu, Z.-Q. (**2007**). "A coupled HMM approach to video-realistic speech animation," Pattern Recogn. Lett. **40**, 2325–2340.

Yamamoto, E., Nakamura, S., and Shikano, K. (**1998**). "Lip movement synthesis from speech based on hidden Markov models," Speech Commun. **26**, 105–115.

Yehia, H., Rubin, P., and Vatikiotis-Bateson, E. (**1998**). "Quantitative association of vocal-tract and facial behavior," Speech Commun. **26**, 23–43.

Yehia, H., Kuratate, T., and Vatikiotis-Bateson, E. (**2001**). "Linking facial animation, head motion and speech acoustics," J. Phonetics **30**, 555–568.

Zelezny, M., Krnoul, Z., Cisar, P., and Matousek, J. (**2006**). "Design, implementation and evaluation of the Czech realistic audio-visual speech synthesis," Signal Process. **86**, 3657–3673.

Perception and production of /r/ allophones improve with hearing from a cochlear implant[a)]

Melanie L. Matthies
Department of Speech Language and Hearing Sciences, Boston University, Boston, MA 02215 and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Frank H. Guenther
Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215-0000 and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Margaret Denny
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Joseph S. Perkell
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215; and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Ellen Burton
Speech Communication Group, Research Laboratory of Electronics,Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Jennell Vick
Department of Speech and Hearing Sciences, University of Washington, and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Harlan Lane
Department of Psychology, Northeastern University, Boston, MA and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Mark Tiede
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Haskins Laboratories, New Haven, CT 06511

Majid Zandipour
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215

(Received 23 January 2008; revised 19 August 2008; accepted 20 August 2008)

Tongue shape can vary greatly for allophones of /r/ produced in different phonetic contexts but the primary acoustic cue used by listeners, lowered F3, remains stable. For the current study, it was hypothesized that auditory feedback maintains the speech motor control mechanisms that are constraining acoustic variability of F3 in /r/; thus the listener's percept remains /r/ despite the range of articulatory configurations employed by the speaker. Given the potential importance of auditory feedback, postlingually deafened speakers should show larger acoustic variation in /r/ allophones than hearing controls, and auditory feedback from a cochlear implant could reduce that variation over time. To test these hypotheses, measures were made of phoneme perception and of production of tokens containing /r/, stop consonants, and /r/+stop clusters in hearing controls and in eight postlingually deafened adults pre- and postimplant. Postimplant, seven of the eight implant speakers did not differ from the control mean. It was also found that implant users' production of stop and stop+/r/ blend improved with time but the measured acoustic contrast between these was still better in the control speakers than for the implant group even after the implant users had experienced a year of improved auditory feedback.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2987427]

PACS number(s): 43.70.Mn, 43.66.Ts [BHS] Pages: 3191–3202

I. INTRODUCTION

Variability in the articulation of /r/ in American English has been studied extensively (see Delattre and Freeman, 1968; Stevens 1998; Westbury *et al.*, 1998; Zhou *et al.*, 2008) with respect to the strong influence of adjoining phonemes and range of individual speaker productions. Large variations in vocal tract shape for /r/ can be observed in different phonetic contexts, but these routinely provide a lowered F3 in the acoustic signal which in turn supplies a reliable perceptual cue to the listener. In discussing the six major types of tongue shapes used in American English, Delattre and Freeman (1968) note, "…the spectrographic pattern is more significantly affected by syllabic position than by tongue shape" (page 52). Variety in vocal tract shape has not been found to distort the resultant acoustic output (for F3 and lower formants) for /r/ produced by normal speakers nor do listeners perceive differences between retroflexed and bunched articulations. Zhou *et al.* (2008) found significant differences between F4/F5 patterns in speakers with bunched versus retroflexed /r/ productions but concluded that these patterns affected the perception of speaker identity rather than phonetic content. With a range of articulatory configurations providing the lowered F3 that corresponds to the /r/ percept, a strong case can be made that the speaker's task is to provide a consistent acoustic cue across a variety of contexts and to implement whatever articulatory configuration will best lower F3 in a given context.

The observation that many articulatory configurations can be found in American English /r/ indicates significant complexity in the production of /r/. Articulatory variability and the influence of context may present difficulties in the acquisition of /r/ by children and second-language learners. In a clinical intervention study, Adler-Bock *et al.* (2007) described /r/ as one of the most complex phones in English because of the multiple articulatory configurations involved in its production. Auditory feedback is a potential mechanism for constraining the acoustic variability of /r/ so that the speaker can provide the listener with a stable cue for perception. The speech motor control system can use auditory feedback in planning the vocal tract shape so that the listener's percept remains /r/ despite a range of articulatory configurations employed by the speaker. The many-to-one relation between articulation and realized /r/ has not been studied extensively in postlingually deafened adults. This is an interesting population because the process of using auditory feedback in lowering F3 will be particularly challenging for a speaker who has difficulty hearing this cue. Consequently /r/ production may degrade in persons with hearing loss.

A. Effects of hearing loss and improved auditory feedback

While speakers with acquired deafness may remain highly intelligible for years after losing hearing, some deterioration in articulatory precision is to be expected (see Lane *et al.* 1995) and certain phonemes may be more susceptible than others. All phonemes can show variation due to coarticulation, but the semivowels are particularly labile with changes in phonetic context. The literature suggests that the production of semivowels is not characterized by strong articulatory anchor points, saturation effects, or robust orosensory feedback. Speakers therefore may rely more on auditory information to guide the production of /r/ than for other consonants. Postlingually deaf speakers may have difficulty in consistently producing lowered F3 and be less able to manage the disparate /r/ articulations induced by varying context. In accord with this idea, some previous work has indicated that deafness may cause a decrease in the /r/-/l/ contrast (Perkell *et al.*, 2001). Because of hearing loss, the auditory feedback of deaf speakers can be insufficient for accurately and reliably mapping between the many articulatory possibilities of /r/ due to context variation and the goal of lowering F3 to produce an intelligible percept for the listener (see Guenther *et al.*, 1998; Guenther *et al.*, 2006; Perkell *et al.*, 2001).

When adventitiously deafened adults receive cochlear implants, they generally show improvements in speech production that are consistent with the objective of transmitting an intelligible message to the listener. Acoustic measures reveal diminished phoneme contrasts in the speech of those deafened adults compared to hearing controls. When some hearing is restored with an implant, gradual improvements are commonly seen (as measured via acoustic analyses and/or intelligibility testing of the speech). Improvements in speech from postlingually deafened adults have been demonstrated longitudinally over several months or even years. In experiments that switch the implant speech processor off and on, phoneme contrasts can improve quickly when auditory feedback is available. These findings apply both to vowel contrasts (Perkell *et al.*, 1992; Lane *et al.*, 2001; Kishon-Rabin *et al.*, 1999) and consonant contrasts (Lane *et al.*, 1995; Matthies *et al.*, 1994, 1996; Waldstein, 1990).

B. A modeling framework

To provide a framework for studying the normal and disordered production of phonemes, models of speech production try to account for the many-to-one relationship between articulation and the acoustic output. In an acoustic and articulatory study of /r/ production in seven speakers with normal hearing, Guenther *et al.* (1999) observed that the participants demonstrated a trade-off between front-cavity length and the size and length of the constriction in different contexts, keeping F3 low—the primary cue for /r/. Using data obtained with an electromagnetic articulometer, Guenther *et al.* (1999) found that productions of /r/ sounds in varying phonetic contexts were acoustically similar despite large differences in their associated articulatory configurations. They noted further that speakers were modifying front-cavity length and constriction length to preserve a total cavity volume that would maintain the acoustic target (low F3) of /r/. A modeling study of articulatory and acoustic variability based on two of the speakers of Guenther *et al.* (1999) supported the idea that the control of /r/ production is based on a low F3 auditory/acoustic target (Nieto-Castanon *et al.*, 2005).

[a] A portion of this work was presented at the 150th meeting of the Acoustical Society of America in Minneapolis, MN, Fall, 2005.

The claim for an auditory goal for /r/ also receives support from a study of native speakers of Japanese learning English as a second language. Such speakers typically have difficulty producing the English /r/-/l/ contrast. Bradlow *et al.* (1999) used a perceptual training paradigm that presented many exemplars to their participants to help them form new phonetic categories, without any production training. They found improvements in production that were correlated with improvements in perception. Along similar lines, Ingram and Park (1998) in a study with Japanese and Korean speakers noted that both pretraining phonological knowledge of the /r/-/l/ contrast and discriminability of the acoustic differences between /r/ and /l/ were important for the identification and discrimination of the phonemes.

C. Study objectives

For the current study, an experiment was designed to examine the role of auditory feedback in the production of /r/ by studying the effect of context on /r/ when produced by postlingually deafened adults. The current study is based on the general hypothesis that speakers attempt to satisfy listener demands for clarity while minimizing their expenditure of effort (Lindblom, 1990; Perkell *et al.* 1997, 2004). Speakers are presumed to have internal models of their speech production mechanisms and acquire acoustic-phonetic knowledge that combine to guide their efforts to produce intelligible speech. In this internal model, movement goals for phonemes can be specified in the articulatory and auditory spaces, as set forth in the Directions Into Velocities of Articulators (DIVA) neurocomputational model of speech motor planning (Guenther, 1995; Perkell *et al.*, 2001). These goals are acquired, maintained, and achieved through the use of feedback and feedforward control mechanisms, which consist of acquired mappings between articulatory and sensory parameters. In particular, there is a primary auditory goal for /r/, which consists mainly of a range of low values of F3. This single auditory goal region can be achieved with many different configurations of the tongue but two, bunched and retroflexed, represent the extremes of this continuum. According to the model, there would be two articulatory-to-auditory mappings for /r/, one for each extreme articulatory configuration. In postlingually deafened adults who have been without hearing for a long time, such mappings are presumed to have degraded somewhat, leading to more variability in the production of auditory/acoustic goals. These ideas led to the formulation of three hypotheses that were tested on two groups of speakers, postlingually deafened adults who received cochlear implants and normal-hearing controls.

The study tests a *context variability hypothesis*: varying phonetic context (as in the nonsense words wabrav, wadrav, wagrav, and warav) will yield more variable /r/ productions by implant users than by controls. The study also evaluates an *allophonic consistency hypothesis*: implant users whose speech perception has improved with experience using their implant will yield more consistent /r/ productions across phoneme context, similar to the pattern exhibited by hearing controls. According to the model, experience using their implants would allow these speakers to recalibrate auditory feedback mechanisms and, using that recalibrated feedback, to modify mappings between disparate articulatory configurations and a more consistent auditory goal for /r/ across phonemes. An additional look at the role of auditory feedback was devised by manipulating the auditory feedback condition for the normal-hearing speakers with masking noise and for the cochlear-implant users by turning their speech processors off during part of the experimental sessions. The *consonant contrast hypothesis* tests that the contrast between stop+r and stop-alone productions is greater for speakers with normal hearing than for implant users. If auditory feedback is useful for maintaining precise articulation of clusters, these contrasts can be expected to vary with feedback state and experience with the cochlear implant.

II. METHOD

A. Participants

The experimental group consisted of six male and two female postlingually deaf, adult, paid volunteers, who received cochlear implants at an average age of 52. The implant was either the Clarion (Advanced Bionics, CIS strategy; Wilson *et al.* 1995) or the Nucleus device (Cochlear Corp.; Blamey *et al.* 1987; McKay and McDermott, 1993). The implant users were referred by the Massachusetts Eye and Ear Infirmary or the University of Massachusetts Memorial Medical Center. The control group consisted of six females and one male; they had no reported difficulties with speech or hearing. All were paid volunteers with an average age of 38, older than the standard college students and therefore are more similar in age to the cochlear-implant users. The implant users were tested three times: preimplant (before their speech processors were turned on), after one month of experience with the implant, and after a year of implant use. Most of the normal-hearing control speakers served for a single session, but several were tested twice to confirm that the intersession variability for these measures was low. Details about the study participants are given in Table I. Each speaker participated in tests of speech production and perception.

B. Speech perception test

For the allophonic consistency hypothesis to be tested, participants with hearing loss must have improved their speech perception through the use of their cochlear implants. Multiple repetitions of the consonants /b, d, g, p, t, k, s, ʃ, z, r, and l/ in a /Cɑd/ environment were recorded by a normal-hearing male speaker. These were presented at a comfortable listening level to the implant users preimplant and at one month and one-year postimplant to measure longitudinal changes in consonant perception. Implant candidates who wore hearing aids preimplant used their aids during the phoneme recognition task in the session prior to implant activation but in the sessions at one month and one-year postactivation, they used only their cochlear implants. Further methodological details and results of the speech perception task have been given elsewhere (Lane *et al.*, 2007).

TABLE I. Participant demographic information, vowel and consonant percent correct identification for implant users, and normal-hearing controls. (Asterisks denote missing data.)

Implant users		FI	FJ	MJ	MK	ML	MM	MO	MP
Age at onset and (duration) profound hearing loss		54 (2)	45 (1)	43 (6)	28 (0)	48 (0)	72 (6)	67 (5)	26 (10)
Cochlear implant		Nucleus 24	Clarion High-Focus	Clarion High-Focus	Clarion Auria	Nucleus 24	Clarion High-Focus	Clarion High-Focus	Clarion High-Focus
Vowel perception (% correct)	Pre	29	26	35	25	64	32	51	27
	1 year	62	55	96	95	66	77	*	74
Consonant perception (%)	Pre	26	25	21	24	21	32	49	21
	1 year	44	57	74	92	82	50	*	74
Normal-hearing controls		FNH1	FNH2	FNH3	FNH4	FNH5	FNH6	FNH7	MNH1
Vowel perception (% correct)		99	98	99	99	98	*	96	97
Consonant perception (%)		91	90	99	99	96	96	98	99

Participants identified the presented consonant, choosing from a full set of 11 /Cɑd/ options displayed on the computer monitor; they used a keyboard or a touch screen to indicate their choice. After a practice run, participants were presented with 12 repetitions of each token arranged in semirandom order, and the responses were recorded and stored for further analysis. The normal-hearing participants were tested once.

C. Speech production elicitation

For each session, the speakers were seated in a sound-attenuating room in a comfortable chair with a head-mounted electret microphone (Audio-Technica, model AT803B) placed at a fixed distance of 20 cm from the lips. Elicitation utterances were randomized prior to the testing session and presented via a computer monitor to prompt the speaker's productions. The microphone output was low pass filtered at 7.2 kHz and digitized in real time at a 16 kHz sampling rate. In each session and condition, speakers produced 15 repetitions of words containing /r/, wabrav, wadrav, wagrav, and warav and the contrasting words without /r/, wabav, wadav, and wagav. The words were embedded in the carrier phrase "Say__for me" (see Boyce and Espy-Wilson, 1997; Guenther *et al.*, 1999).

In the experimental condition in which feedback was blocked (feedback off), participants with normal hearing received masking noise at 95 dB sound pressure level (SPL) through calibrated TDH-39 headphones. The noise was generated by a custom-built device (Technical Collaborative, Lexington, MA) and was approximately speech shaped, with a spectral envelope that rolled off at 6 dB per octave. The implant users kept the normal settings of their speech processors for the feedback-on conditions and turned off their processors for the feedback-off conditions. For the first half of the experiment, speakers had full access to auditory feedback (feedback on). For the second half of the experiment, implant users turned off their speech processors while participants with normal hearing received speech-shaped noise during their productions (feedback off). With seven test words and 15 repetitions of each, there were 105 utterances in each of the two listening conditions. Test words were randomized within each condition.

1. Acoustic analysis

As schematized in Fig. 1, the first three formant frequencies were extracted algorithmically from a linear predictive coding (LPC) spectrum at the consonant-/r/ boundary (dotted vertical line) and at the midpoint of the following /ɑ/. The LPC filter order was chosen to optimize formant extraction

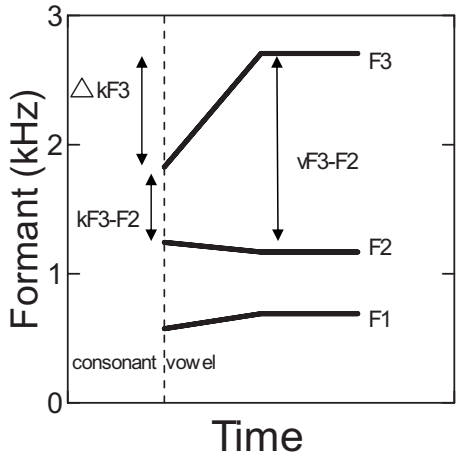


FIG. 1. Schematized formant tracks that depict measurements taken during the vowel, from the consonant-vowel boundary (dotted line) through the midvowel, and how these were used for analyses.

for each speaker. Initially, the extraction algorithm displays a broadband spectrogram with superimposed formant trajectories. If the three formants at the midvowel point were detected in frequency regions expected for /ɑ/, the experimenter accepted that token with those values as complete. The analysis parameters (LPC order, window size, etc.) had gender-specific default settings, but the experimenter could adjust these parameters for individual speakers. If the algorithm did not succeed in detecting all of the necessary formants, the experimenter could vary the measurement time slightly. A small shift in the analysis time (to a different phase in the glottal cycle) would usually result in usable formant values. The detected formant values could also be modified manually if a formant peak was missing because of excessive bandwidth or peak merging. Such corrections were guided by observation of the concurrently displayed discrete Fourier transform (DFT) spectrum. In general, the articulation of /r/ can be classified as a lingua-palatal glide or semivowel (Ong and Stone, 1998). In accordance with this idea, the analysis extracted values from the formant trajectories according to the recognition algorithm of Espy-Wilson (1992) and theory described in Stevens (1998).

Six measurements were obtained from the formant tracks: F1, F2, and F3 at the consonant-vowel boundary (vertical dotted line) and F1, F2, and F3 at the midvowel. Analyzed variables included F1 and F2 at the midvowel, the difference between F2 and F3 at the consonant-vowel boundary (kF3−F2) and at the midvowel (vF3−F2), and the change in F3 (ΔkF3) through the vowel. The arrow labeled ΔkF3 in Fig. 1 designates the change in F3 (i.e., the F3 transition) from its minimum at or near the consonant-vowel boundary (dotted vertical line) to the middle of the vowel. The consonant-vowel boundary alone was used for the singleton consonant tokens (wabav, wadav, and wagav) without regard to whether F3 had reached a minimum (or maximum) but the minimum F3 was used for all /r/ tokens. Thus the extraction and analysis of the acoustic data were as parallel as possible across token types. The midvowel formants for /ɑ/ were expected to be fairly similar across contexts while the formant transitions from the consonant-vowel boundary were expected to be strongly influenced by the specific consonant environment. In general, /r/ is characterized by a small initial difference between F3 and F2 (kF3−F2) with a sharp rise in F3 by midvowel (vF3−F2). McGowan *et al.* (2004) successfully used these formant-based measures in their study of /r/ production and found them to be robust across extensive morphological differences among their study participants, who were as young as 14 months of age.

2. Estimating variability of /r/ relative to /ɑ/

To illustrate the variability phenomenon of interest in /r/ production, consider the disparate frequency distributions of the excursion of F3 transitions for /gr/ and /dr/ productions from a single speaker with hearing loss in the left half of Fig. 2. The distributions of the acoustic measure from a retroflexed /r/ produced in wadrav compared to those from a bunched /r/ produced in wagrav are expected to be distinct from one another if the speaker did not have well-tuned map-

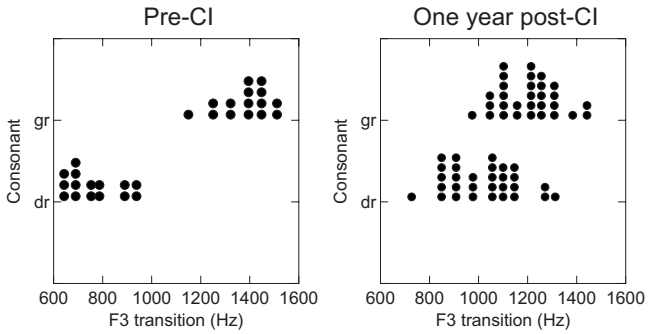


FIG. 2. A comparison of F3 transition data for wadrav (likely to be retroflexed) and wagrav (likely to be bunched) as a function of experience with the cochlear implant for participant ML.

pings between the two different articulatory configurations and a single auditory goal for /r/. Indeed this was clearly the case for the implant user whose data are shown in Fig. 2 when acoustic parameters of /r/ were measured preimplant. In this speaker, considerable allophonic variation between /dr/ and /gr/ productions was seen in the preimplant productions (left panel); however, the distribution of the F3 transitions showed a postimplant pattern (right panel) consistent with normal hearing. That is, the distributions largely overlap in the one-year session (which includes both feedback conditions, right-hand panel), indicating less contextual influence on the acoustic realization of /r/. Other speakers showed this increase in allophonic overlap for the F3 transition but to a more limited extent, possibly due to the range and variety of articulation strategies that speakers use to produce /r/ (see Guenther *et al.*, 1999).

To provide a metric that was robust across all speakers, a ratio of /r/ dispersion to /ɑ/ dispersion was constructed. As shown in Fig. 3, variability for /ɑ/ was estimated as the dispersion of midvowel values of F1 and F2 for individual tokens around the center of their distribution (shown by arrows in the figure) in the F1, F2 plane. Dispersion was calculated as the Euclidean distance of each token from the distribution mean, averaged across all wabav, wadav, and wagav *tokens*. Dispersion was calculated for /r/ in a parallel way using values in the plane defined by the F3 transition (in hertz) and the F3−F2 difference (in hertz) at the /r/-vowel boundary.

A variability index was defined as the ratio of /r/ dispersion (average of distances to the /r/ center) to /ɑ/ dispersion (distances to the vowel center). To calculate the variability index, tokens were matched pairwise by repetition number for wabrav to wabav, wadrav to wadav, and wagrav to wagav. Pairing was used to provide approximate comparability for the dispersion comparisons with respect to potential warm-up and fatigue effects across the length of the experiment. For each pair, the /r/ distance to center was divided by the /ɑ/ distance to the center. The variability index is the average of these ratios across multiple token pairs; it is closer to 1.0 when the dispersion measure for /r/ was more similar to the dispersion for /ɑ/ as found in the bottom panel. When /r/ variability is much greater than vowel variability, the variability index value is much greater than 1.0, as seen in the top panel of Fig. 3. When the allophonic variability is reduced, /r/ dispersion also decreases while the /ɑ/-dispersion

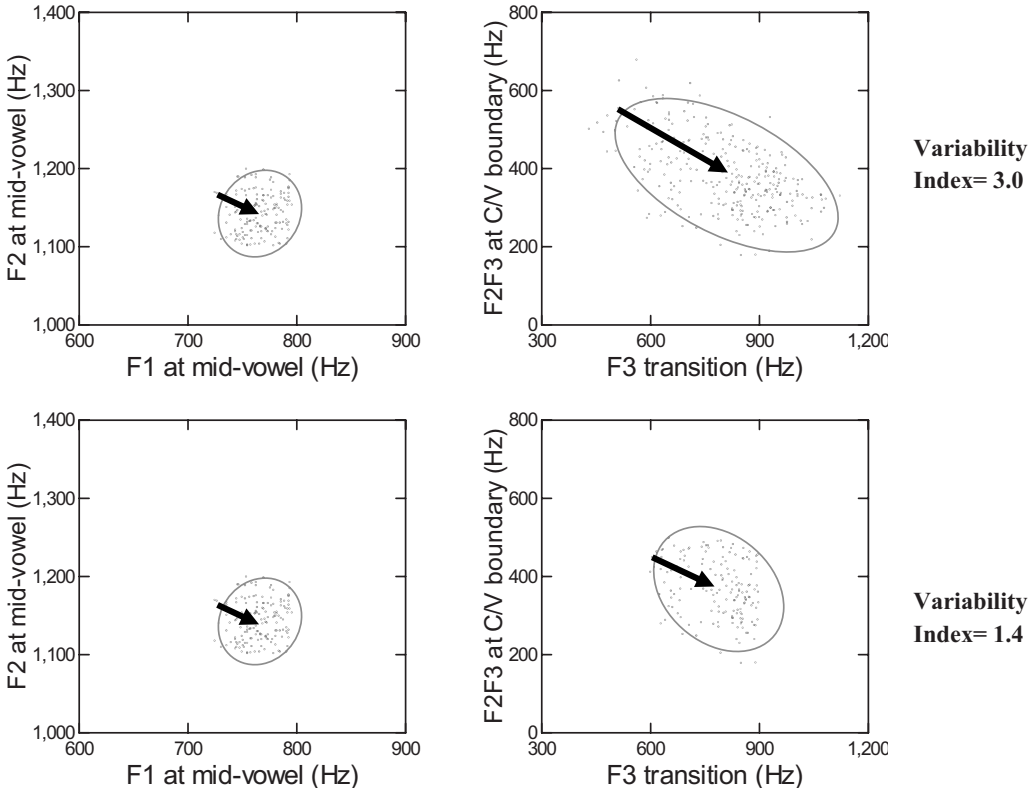


FIG. 3. Schematized comparison of a small variability index (lower panels) with a larger variability index (upper panels) due to greater /r/ dispersion in the upper right-hand plot. The /r/ dispersion is plotted in the domain of the F3−F2 difference at the beginning of the vowel by the F3 transition from vowel beginning to midvowel. The vowel dispersion is plotted in the F1, F2 domain.

remains roughly constant, yielding smaller variability indices (lower panels in Fig. 3). Log-based dispersion metrics were found to have a more normal distribution than linear ones, so logarithmically transformed measures were used for statistical analyses.

III. RESULTS

A. Speech perception

Vowel- and consonant-identification results for the normal-hearing participants are presented in the last two rows of Table I; they show very high or perfect scores. Table I also illustrates considerable gains in vowel and overall consonant identification for the implant users from the preimplant condition to one-year postimplant. As shown in Fig. 4, the implant users' overall proportion correct for consonant identification rises from less than 30% in the pre-cochlear implant (CI) condition to nearly 70% at the one-year session. Overall consonant identification and identification of the semivowels, voiced stops, and sibilants progress most rapidly between the preimplant condition and one-month testing period. Paired *t*-tests for /r/, sibilants, and /b, d, g/ identification revealed significant differences between pre- and one-month results for all three consonant types but no statistically reliable differences between the one month and the one-year sessions. The perceptual improvements noted in Table I and Fig. 4 support the goal of testing the allophonic consistency hypothesis to determine if corresponding speech production benefits could be observed when comparing preimplant to postimplant speech.

B. Speech production

The speech production results are presented in three sections. The first section reports findings from between-group comparisons to assess the context variability hypothesis, which is concerned with the variability of /r/ allophones for the implant group as compared to the speakers with normal hearing. The second section presents longitudinal results to evaluate the allophonic consistency hypothesis, which predicts that implant users will show a reduction over time in /r/ variability across phoneme context, making their results approach those from controls. The third section compares con-

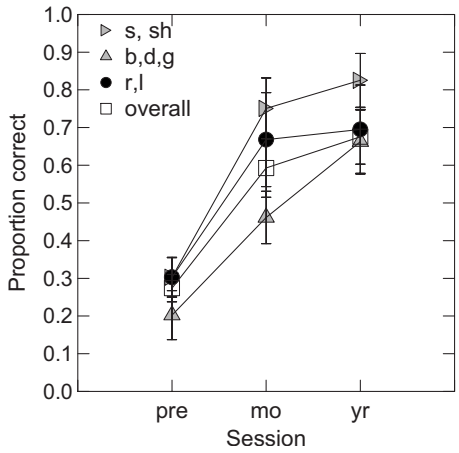


FIG. 4. Proportion correct from 11-item consonant confusion matrices constructed from results of consonant identification (12 repetitions per consonant) at a comfortable listening level.

trasts implemented between singleton stops and stop+/r/ in the two groups, thereby addressing the consonant contrast hypothesis, which states that those contrasts will be greater for speakers with normal hearing. The effects of experimentally induced auditory feedback changes (by adding masking noise) are included to provide a condition with greater difficulty (and therefore more variability) in the results for the normal-hearing speakers which may in turn provide a more realistic comparison for the results from the cochlear-implant users.

1. Between-group comparisons for variability analyses

A variability index was defined for each speaker as the average ratio of /r/ dispersion (average distance to the /r/ distribution center) to /ɑ/ dispersion (average distance from their vowel mean), as explained in Sec. II. This ratio, rather than /r/ variability alone, was used to reduce between-speaker variability, particularly the potential effect of gender. Group averages for the implant users and normal-hearing participants were calculated by pooling values from individual token contrastive pairs across all the participants in each group. Although measures such as F0 and the formants in these words clearly varied with gender, the frequency distributions of variability indices for the two female implant users were statistically indistinguishable from those of the male participants.

The mean variability index for the control speakers with normal hearing was 1.22 (sd=0.35); there was greater variability in /r/ dispersion than in /ɑ/ dispersion. When auditory feedback was disrupted with masking noise (feedback off) in normal-hearing controls, larger variability indices (mean =1.33, sd=0.40) were found ($F=13.9$, $p<0.001$, and effect[1] size $\eta p^2=0.12$). The main effect for plosive place of articulation was also significant ($F=14.5$, $p<0.001$, $\eta p^2=0.13$) with /br/ comparisons to /ɑ/ in the feedback-on condition having the smallest variability index (mean=1.08) and /gr/ comparisons to /ɑ/ having the largest (1.35).

For the speakers with cochlear implants, repeated-measures ANOVAs were conducted with sessions, consonant place, and feedback as conditions, speaker as a category variable, and variability index as the dependent variable. The preimplant data revealed a significant main effect for participant ($F=6.32$, $p<0.001$, $\eta p^2=0.29$) but consonant place (br, dr, gr) was not significant nor was there an interaction between place and speaker. Consistent with the context variability hypothesis, all of the hearing-impaired participants had greater preimplant variability indices (mean=1.43, sd =0.43) than the normal-hearing controls (mean 1.22, as reported above) with feedback on, although one (FJ) had a variability index distribution similar to the controls (one-sample t-test $p>0.05$). Four of the implant users' preimplant variability index values approached the variability index mean of the feedback-off performance for normal-hearing controls (1.33, as reported above) while three other implant users yielded variability index values higher than those found in either normal hearing (NH) condition. (One implant user (MM) could not be included in the analysis because preimplant data were not collected).

TABLE II. Effect sizes (ηp^2) from the three ANOVAs.

ANOVA effect	NH	CI month	CI year
Speaker	0.43	0.26	0.44
Place	0.13	0.06	0.06
Place X speaker	0.09	0.14	0.13
Feedback	0.12	0.07	0.01
Feedback X speaker	0.43	0.11	0.21

A summary of speaker, place, and feedback effect sizes (partial eta squared, ηp^2) from the ANOVAs for controls and implant users (at two time intervals) is found in Table II. There are substantial differences between the measures for implant users and those for speakers with normal hearing. The speaker main effect for variability index analyses was statistically significant in all three ANOVAs (NH, CI month, and CI year) as was the feedback-by-speaker interaction. Feedback and place (consonant context) were less predictive of allophonic variability in the implant users than in the normal-hearing controls, and the interaction of place-by-speaker was larger for the implant participants, suggesting that certain clusters were more difficult for some implant participants to produce consistently. To summarize, variability indices were greater than 1.0 for all speakers; that is, /r/ was more variable than /ɑ/ under contextual variation. The context variability hypothesis was supported in that varying phonetic context (wabrav, wadrav, wagrav, and warav) yielded more variable acoustics for /r/ from implant users than controls.

2. Longitudinal results

The changes in implant-users' variability indices as they gained experience with prosthetic hearing were analyzed to test the allophonic consistency hypothesis, which postulated that /r/ variability would become more similar to that of controls (i.e., would diminish) over time with the use of the implant. All of the implant participants decreased their variability across sessions so that in the one-month session, one-year session, or both, it was more similar to the results from participants with normal-hearing listening to masking noise in the feedback-off condition. There was a significant main effect of session for implant users during the feedback-off condition ($F=9.9$, $p<0.001$, $\eta p^2=0.07$). There was surprisingly no significant main effect of session during the feedback-on condition, a result that will be explored further below.

a. Individual speaker differences. Despite the indications that prosthetic hearing reduced allophonic variability of /r/ for the group, differences among speakers complicate the picture. The ANOVA results reveal considerable individual variability in the performance of cochlear-implant users across sessions in general and in their reduction in context-induced variability with prosthetic hearing. There was a significant speaker main effect for variability index (feedback on: $F=11.06$, $p<0.001$, $\eta p^2=0.45$ and feedback off: $F=9.64$, $p<0.001$, $\eta p^2=0.40$) and an interaction between speaker and session (feedback on: $F=3.23$, $p<0.001$, $\eta p^2=0.19$ and feedback off: $F=3.35$, $p<0.001$, $\eta p^2=0.19$). The

consonant place by speaker interaction was statistically significant for both the feedback-on ($F=2.035$, $p=0.024$, $\eta p^2=0.13$) and feedback-off ($F=2.42$, $p=0.006$, $\eta p^2=0.14$) analyses.

Many of the implant users produced some tokens in which the context-induced variability of /r/ was much greater than the variability of /ɑ/ (resulting in variability indices ≥ 2.0) even in the one-year session. Thus, these implant users had variability index values that greatly exceeded those for normal-hearing controls when the latter had access to auditory feedback. A few implant users, however, had some variability index values in postimplant sessions that were lower than the normal-hearing controls' mean. The session by speaker interaction was significant ($\eta p^2=0.19$) for both feedback conditions, indicating that some speakers showed a consistent longitudinal trend toward reduced /r/ allophonic variability but others did not.

b. Feedback effects. Overall, three implant users yielded relatively small variability index values and near-normal performance in both feedback conditions by the end of the study. Four implant users showed decreased variability indices in at least one feedback condition as the study progressed. Compared to the preimplant measures, postimplant variability indices showed improvement even when the implant was turned off (feedback-off) as evidenced by a significant session effect ($F=6.91$, $p=0.001$, $\eta p^2=0.07$). The main effect for speakers was more pronounced with auditory feedback available from the implant (feedback-on) condition, and the corresponding session effect was not significant ($F=0.62$, $p=0.539$). When the distribution of differences across repetitions of the matched pairs was examined, there were two implant participants in the feedback-on condition who had distributions similar to those obtained in the normal-hearing (feedback-on) condition.

Two observations lead to the inference that the implant users developed stable articulatory targets that were not degraded when their implants were turned off. First, there was a significant main effect of session for implant users during the off condition ($F=9.9$, $p=0.001$, $\eta p^2=0.07$). Hence, the variability index improved over the three sessions even in the absence of auditory feedback. Second, the feedback main effect was statistically significant only in the one-month session ($F=7.3$, $p<0.01$) and even there did not account for much variance ($\eta p^2=0.07$). Thus, interrupting feedback did not appreciably impair performance. Continued improvements in the variability index in postimplant sessions, therefore, were mostly independent of implant feedback state during testing. Conclusions about group longitudinal trends in variability must be tempered because of the various interaction terms involving participants. In particular, there was no significant effect of session during the feedback-on condition ($\eta p^2=0.01$) due to a heightened speaker main effect. In general, however, a decrease in variability index toward normal values was observed for many of the cochlear-implant users, supporting the allophonic consistency hypothesis.

3. Contrasting stop+r with stop alone

To assess changes in the contrast between a singleton stop and that stop followed by /r/ (stop+/r/) and test the consonant contrast hypothesis (i.e., contrast will be greater for normal-hearing control speakers than for implant users), the distinctiveness of /b/, /d/, and /g/ in wabav, wadav, and wagav was measured using the magnitude of the F3 transition (ΔF3) and the F3−F2 (kF3−F2) difference to classify tokens of the phonemes produced.

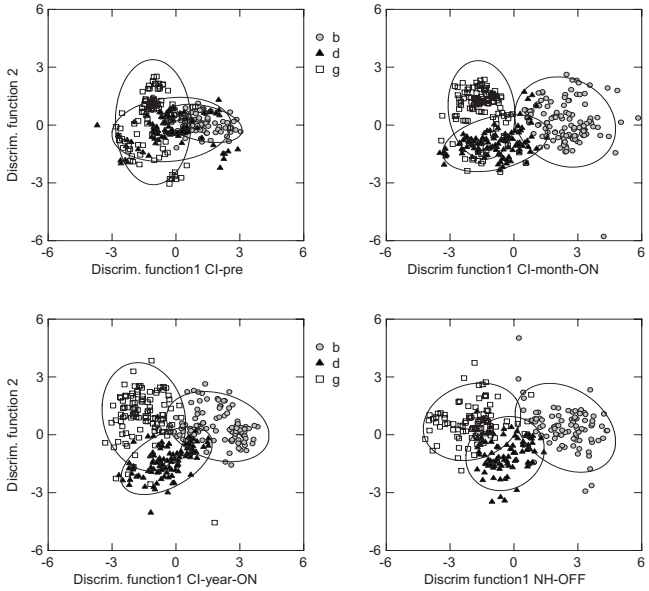


FIG. 5. Results of discriminant analyses of speakers' productions of /b/, /d/, and /g/.

a. Classification analysis with normal-hearing controls. The lower right-hand panel of Fig. 5 shows the discriminant analysis classification (90% correct) with the three stop phonemes for the normal-hearing controls in the feedback-off (95 dB SPL masking noise) condition. The less-than-perfect classification was not attributable to a lack of intelligibility in the stop productions by the normal-hearing controls. Rather, it occurred because F2 and F3 measures did not capture all of the crucial information in this contrast. Although adding information about F1 transitions would greatly improve the discriminant analysis results by increasing the separation among /b/, /d/, and /g/, the current analysis focused on F2 and F3 so that the singleton stops could be compared with stop+/r/ clusters.

b. Classification analysis with implant users. The upper left-hand panel of Fig. 5 shows the discriminant-analysis classification of the stops produced by the postlingually deaf speakers prior to experience with the cochlear implant. There was considerable overlap among the consonant groups, with 66% of the tokens classified correctly. While /b/ tokens (lightly shaded circles) were tightly clustered and distinct from /g/ (open squares), the /d/ productions (filled triangles) were quite variable. In the one-month data (upper right-hand panel) and one-year data (lower left-hand panel) it can be seen that the result for the discriminant analysis of stop productions by implant users with feedback-on and a year's experience with their implants (89% correct classification) is comparable to that of the normal-hearing controls (90%—lower right-hand panel). The most striking improvement was seen for /d/, which was classified correctly 42% of the time in the preimplant condition but increased to 92% correct at

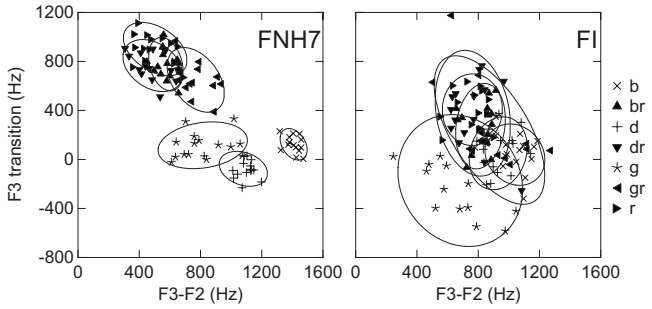


FIG. 6. A comparison of singleton stops (open symbols) vs stop+r tokens (closed, filled symbols) for a normal-hearing speaker with auditory feedback (left panel) and a speaker with hearing loss in the preimplant condition (right panel).

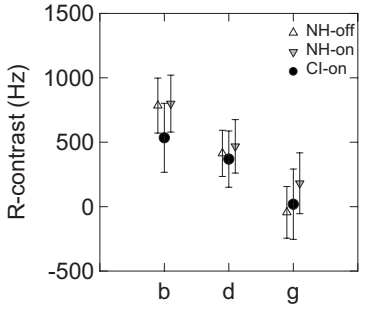


FIG. 7. Extent of /r/ contrast measured as the difference in the F3 transition for singleton stop vs stop+r productions with the indicated stops. Results are shown for implant users in the one-year session (filled circles) and for normal-hearing speakers with their hearing masked (clear triangles) or not masked (filled triangles).

the one-month session and remained relatively distinct from the other two consonants in the one-year analysis.

c. Between-group comparisons and formant measures. Normal-hearing speakers with better perceptual acuity have been found to produce more distinctive vowel contrasts (Perkell *et al.*, 2004). A listener must perceive certain specific cues to differentiate between a speaker's productions, for example, of wabav and wabrav, so the purpose of these analyses was to evaluate the contrast between the stop productions and their stop+/r/ counterparts under varying feedback conditions. As noted previously, singleton stops can be classified using parameters based on F2 and F3. Figure 6 plots the extent of the F3 transition (see Fig. 1: ΔkF3) against the F3−F2 difference (see Fig. 1: kF3−F2) at the beginning of /ɑ/ for a normal-hearing speaker (left-hand panel) with auditory feedback. Tokens for the three plosives (stop+r, filled symbols) overlap with each other (and also with /r/ alone from the warav tokens). When these stops are followed by /r/, they are quite divergent from the singleton stops without /r/. Normal-hearing controls thus reveal a contrast between the stops and the stop+r counterparts with respect to the F3 transition and the difference between F3 and F2 at the consonant/vowel boundary. Before receiving their implants, hearing-impaired speakers generally produced less contrast between stops and their stop+r counterparts; a typical result is seen in the right-hand panel of Fig. 6.

The consonant contrast hypothesis stated that acoustic measures of the difference between stop+/r/ and singleton stop productions will be greater for control speakers with normal hearing than for implant users. The advantage provided by auditory feedback is expected to yield better phonemic contrast among the stops and between each stop and its stop+/r/ counterpart. Figure 7 plots the extent of /r/ contrast measured as the difference in F3 transition for singleton stop versus stop+/r/ productions with the indicated stops. Results are shown for implant users in the one-year session (filled circles) and for normal-hearing controls with their hearing masked (feedback off as clear triangles) or not masked (feedback on as filled triangles). Pooled across the stops, normal-hearing speakers had higher /r/-contrast values (mean=482.1 sd=335.2) than implant users at one year (mean=300.9 sd=339.9) with feedback on (t for independent samples=6.94, $p<0.001$, $d=0.53$).

Inspection of Fig. 7 indicates that the normal-hearing controls, both with and without auditory feedback, as well as the implant users yielded greater contrast for the plosives /b/ and /d/ than for /g/. With the relatively high visibility of the bilabials and the fact that the bilabial+/r/ blend involves two different primary articulations rather than two overlapping gestures of the tongue (as needed for the velar or alveolar stop+/r/), deaf participants may have less difficulty contrasting productions of bilabial+/r/ with bilabial singleton stops.

d. Contrast improvement. Because the preimplant data were obtained with only one feedback condition while the data one month and one year postimplant had two feedback conditions, an ANOVA with session by feedback was not computed. Instead, two repeated-measures ANOVAs were completed. The first had trial variables preimplant, one-month feedback on, and one-year feedback on. The second ANOVA had conditions preimplant, one-month feedback off, and one-year feedback off. There was a significant main effect of session (preimplant, one month, and one yr) on the /r/ versus singleton stop contrast in both the feedback-on ($F=13.17$, $p<0.001$, $\eta p^2=0.14$) and feedback-off ($F=10.83$, $p<0.001$, $\eta p^2=0.11$) conditions. Implant users, therefore, reliably improved contrast over time in both feedback conditions. Implant users varied significantly by session which yielded the largest main effects in these analyses with $\eta p^2=0.42$ for feedback-on and $\eta p^2=0.48$ for feedback-off conditions. As illustrated in Fig. 7, when compared to normal-hearing controls exposed to masking noise (feedback off), implant users have poorer /r/ contrast than participants with normal hearing for /b-br/ and /g-gr/ even with auditory feedback from the implant and one year of implant experience. The /d-dr/ contrast is similar for both groups of participants and for both feedback conditions but the contrast was smallest for /g-gr/ for all of the participants even speakers with normal hearing and optimal access to auditory feedback cues.

The speakers with normal hearing produced greater stop +/r/ versus singleton stop contrast with auditory feedback (Fig. 6, filled triangles, $F=81.04$, $p<0.001$, $\eta p^2=0.45$) than without (unfilled triangles) primarily in the /g/ context. For the implant users, on the contrary, the presence versus absence of feedback did not yield a significant main effect during the one-month ($\eta p^2=0.009$) or one-year sessions (ηp^2

=0.004), although it did interact significantly with the speaker variable (ηp^2=0.11 and 0.29, respectively).

Thus, the contrast between stop+/r/ and stop alone was greater for control speakers with normal hearing than for implant users even when the latter had a year of experience with their implant, although they did improve over the three sessions. Blocking feedback reduced the normal-hearing controls' stop+/r/ contrasts but not those of the implant users.

IV. DISCUSSION

The purpose of the current study was to examine context-dependent allophonic variation in the production of /r/, how it changed longitudinally in cochlear-implant users and how the implant users' production variability compared with that of normal-hearing speakers. This work was intended to address two distinct issues. The first issue is concerned with the nature of motor programming goals in speech production. The second is concerned with the learning that underlies longitudinal reduction in potentially undesirable allophonic variation while maintaining crucial phonemic contrasts. These two issues are joined by a theoretical view of the role of auditory feedback in speech production and how individuals use feedback from a cochlear implant.

A. Goals in a model of speech production

It has been suggested that speakers attempt to satisfy listener demands for clarity but are constrained by economy of effort (Lindblom, 1990; Perkell *et al.* 1997, 2004). As was noted in the descriptions of data from normal-hearing controls, speakers may exhibit a trade-off between clarity and economy of effort and thereby vary widely in the expression of phonetic contrast in their productions. Since /r/ is produced with much greater articulatory variability than almost any other phoneme in English, including the vowel /ɑ/, it was expected and confirmed that the context-induced acoustic variability for /r/ productions would be larger than for /ɑ/ productions for all speakers. This outcome furthered the testing of the context variability hypothesis, which states that varying phonetic context (as in the nonsense words wabrav, wadrav, wagrav, and warav) will yield more variable /r/ productions by implant users than by controls. In their preimplant recording sessions, all of the hearing-impaired participants had greater variability indices than the normal-hearing controls with access to auditory feedback; this finding supports the context variability hypothesis.

Because of the multiple articulatory configurations speakers use in producing /r/ (see Delattre and Freeman, 1968; Guenther *et al.*, 1999; Westbury *et al.*, 1998) and because of relatively crude and frequency-shifted acoustic cues provided by a cochlear prosthesis, implant users may have difficulty in modifying their phoneme-to-articulation maps to produce /r/ appropriately by using their prosthetic hearing. Consequently, implant users would have more difficulty producing semivowels correctly than phonemes with more robust orosensory correlates such as fricative and stop consonants. The consonant contrast hypothesis, which was confirmed, stated that acoustic measures of the difference between stop+/r/ and singleton stop productions will be greater for speakers with normal hearing than for implant users. Greater contrast for the stop+/r/ blends versus the singleton stops (e.g., b versus br, d versus dr, g versus gr) was found for normal-hearing controls with auditory feedback than for implant users with feedback in all sessions.

B. Longitudinal improvements with implant use

The allophonic consistency hypothesis stated that implant users' acoustic measures of /r/ production after experience with their implants will show a reduction in variability and thereby be more similar to those of controls. Indeed, implant users yielded decreasing values of the variability index as their experience with their implants increased. With just a month's experience with their implants, two of the speakers yielded variability index values comparable to those of the normal-hearing speakers. With a year of prosthetic hearing, seven implant users did so.

Perkell *et al.* (2001) discuss the improved production of /r/ by cochlear-implant speakers, noting that acoustic measures of /r/ and /l/ may overlap in preimplant recordings, but the distributions of the two phonemes become distinct and separate after a year of implant use. The participants described in that study were profoundly deaf and did not have any open-set consonant recognition in an auditory-alone task. By contrast, the speakers in the current study had more residual hearing preimplant and their speech did not seem to have deteriorated as much as had been observed in earlier studies. In fact, the productions of /r/ and /l/ preimplant (analyzed with a similar combination of ΔF3 and kF3−F2 measures) in the current implant group were similar to the early postimplant findings of the previous generation of participants. Thus, the acoustic analyses for the current group of speakers had to focus on more subtle acoustic cues than ones that would simply differentiate consonants from one another. To this end, relative acoustic variability, separation among stops in F2−F3 space, and the contrast between singleton stops and stop+/r/ clusters were investigated.

The overall aim of a speaker, with or without a hearing loss, is to present the listener with intelligible speech. The information provided by a cochlear implant can help to refine the planning of speech movements so that the speaker can achieve this goal. In the case of the semivowel /r/, if postlingually deafened adults are not able to produce consistently clear productions across phonetic contexts, they are able to improve clarity once they begin receiving acoustic information from a cochlear implant about their own productions and those of others. In terms of the present theoretical framework, profound postlingual hearing loss likely leads to some gradual degradation of sound contrasts and mappings between articulatory and auditory parameters. In the case of American English /r/, such degradation would cause increased variability of F3 and increased separation of mean values of F3 produced by bunched versus retroflexed articulations. When hearing is partially restored with a cochlear implant, the new unfamiliar auditory stimulation could lead

to some increased variability; however, as mappings and auditory goal regions become refined with time, overall variability and separation of the two allophones decrease.

V. SUMMARY AND CONCLUSIONS

Varying phonetic context (as in the nonsense words, wabrav, wadrav, wagrav, and warav) induced more variability in /r/ productions among implant users than among normal-hearing speakers (supporting the context variability hypothesis). When postlingually deaf speakers were provided some auditory feedback with a cochlear implant, their /r/ productions became more similar to those by hearing controls. Seven out of the eight implant users showed reductions in an index of acoustic variability for /r/ to within normal limits from a preimplant recording to one made one-year postimplant (supporting the allophonic consistency hypothesis). The contrast between stop+r and stop-alone productions was found to be generally larger in speakers with normal hearing than in implant users (supporting the consonant contrast hypothesis). Although implant users showed improvements in the perception and production of stop consonants as well as stop+r blends, their contrast between /b/ and /br/ and between /g/ and /gr/ was still reduced compared to controls, even at one-year postimplant with auditory feedback available.

ACKNOWLEDGMENTS

This research was supported by Grant No. R01-DC003007 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We are grateful to Dr. Donald Eddington of the Massachusetts Eye and Ear Infirmary and Dr. Daniel Lee of the Department of Otolaryngology-Head and Neck Surgery, University of Massachusetts Medical Center for referring cochlear-implant users to our study. We also wish to thank Advanced Bionics and Cochlear for donating research devices for our use.

[1]Partial eta squared, ηp^2, was used as an effect size measure to assess the strength of the association between the sample main effect and the dependent variable. It is calculated as the proportion of the sums of squares due to the effect divided by the sums of squares due to the effect plus the sums of squares due to error (Tabachnik and Fidell, 1996). As partial eta squared approaches 1.0, it reflects more variance accounted for by the effect under study.

Adler-Bock, M., Bernhardt, B. M., and Bacsfalvi, P. (**2007**). "The use of ultrasound in remediation of North American English /r/ in two adolescents," Am. J. Speech Lang. Pathol. **16**, 128–139.

Blamey, P. J., Dowell, R. C., Brown, A. M., Clark, G. M., and Seligman, P. M. (**1987**). "Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors," J. Acoust. Soc. Am. **82**, 48–57.

Boyce, S. and Espy-Wilson, C. Y. (**1997**). "Coarticulatory stability in American English /r/," J. Acoust. Soc. Am. **101**, 3741–3753.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1999**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech perception," J. Acoust. Soc. Am. **101**, 2299–2310.

Delattre, P., and Freeman, D. (**1968**). "A dialect study of American R's by x-ray motion picture," Language **44**, 29–68.

Espy-Wilson, C. Y. (**1992**). "Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English," J. Acoust. Soc. Am. **92**, 736–757.

Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation and rate effects in a neural network model of speech production," Psychol. Rev. **102**, 594–621.

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., and Perkell, J. (**1999**). "Articulatory tradeoffs reduce acoustic variability during American English /r/ production," J. Acoust. Soc. Am. **105**, 2854–2865.

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang **96**, 280–301.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). "A theoretical investigation of reference frames for the planning of speech movements," Psychol. Rev. **105**, 611–633.

Ingram, J. C. L., and Park, S.-G. (**1998**). "Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners," J. Acoust. Soc. Am. **103**, 1161–1174.

Kishon-Rabin, L., Taitelbaum, R., Tobin, Y., and Hildesheimer, M. (**1999**). "The effects of partially restored hearing on the speech production of postlingually deafened adults with multichannel cochlear implants," J. Acoust. Soc. Am. **106**, 2843–2857.

Lane, H., Matthies, M., Denny, M., Guenther, F., Perkell, J., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (**2007**). "Effects of short- and long-term changes in auditory feedback on vowel and sibilant contrasts," J. Speech Lang. Hear. Res. **50**, 913–927.

Lane, H., Matthies, M., Perkell, J., Vick, J., and Zandipour, M. (**2001**). "The effects of changes in hearing status in cochlear implant users on the acoustic vowel space and coarticulation," J. Speech Lang. Hear. Res. **44**, 552–563.

Lane, H., Wozniak, J., Matthies, M. L., Svirsky, M. A., and Perkell, J. S. (**1995**). "Phonemic resetting versus postural adjustments in the speech of cochlear implant users: An exploration of voice-onset time," J. Acoust. Soc. Am. **98**, 3096–3106.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H and H is theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 403–439.

Matthies, M. L., Svirsky, M. A., Lane, H., and Perkell, J. S. (**1994**). "A preliminary study of the effects of cochlear implants on the production of sibilants," J. Acoust. Soc. Am. **96**, 1367–1373.

Matthies, M. L., Svirsky, M. A., Perkell, J. S., and Lane, H. (**1996**). "Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant," J. Speech Hear. Res. **39**, 936–946.

McGowan, R. S., Nittrouer, S., and Manning, C. J. (**2004**). "Development of [ɹ] in young, Midwestern, American children," J. Acoust. Soc. Am. **115**, 871–884.

McKay, C. M., and McDermott, H. J. (**1993**). "Perceptual performance of subjects with cochlear implants using the spectral maxima sound processor (SMSP) and the mini speech processor (MSP)," Ear Hear. **14**, 350–367.

Nieto-Castanon, A., Guenther, F. G., Perkell, J. S., and Curtin, H. D. (**2005**) "A modeling investigation of articulatory variability and acoustic stability during American English /r/ production," J. Acoust. Soc. Am. **117**, 3196–3212.

Ong, D., and Stone, M. (**1998**). "Three-dimensional vocal tract shapes in /r/ and /l/: A study of MRI, ultrasound, electropalatography, and acoustics," Phonoscope **1**, 1–13.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M., Perrier, P., Vick, J., and Zandipour, M. (**2001**). "Planning and auditory feedback in speech production," *Speech Motor Control in Normal and Disordered Speech: Proceedings of the 4th International Speech Motor Conference* (Nigmegen: The Netherlands), pp. 5–11.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (**2004**). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," J. Acoust. Soc. Am. **116**, 2338–2344.

Perkell, J. S., Lane, H., Svirsky, M. A., and Webster, J. (**1992**). "Speech of cochlear implant patients: A longitudinal study of vowel production," J. Acoust. Soc. Am. **91**, 2961–2978.

Perkell, J. S., Matthies, M. L., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J. F., and Guiod, P. (**1997**). "Speech motor control: Acoustic goals saturation effects, auditory feedback, and internal models," Speech Commun. **22**, 227–250.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Tabachnik, B. G., and Fidell, L. S. (**1996**) *Using Multivariate Statistics* 3rd ed. (HarperCollins College Publishers, New York).
Waldstein, R. (**1990**). "Effects of postlingual deafness on speech production: Implications for the role of auditory feedback," J. Acoust. Soc. Am. **88**, 2099–2114.
Westbury, J. R., Hashi, M., and Lindstrom, M. (**1998**). "Differences among speakers in lingual articulation for American English /r/," Speech Commun. **26**, 203–226.
Wilson, B., Lawson, D., Zerbi, M., Finley, C., and Wolford, R. (**1995**). "New processing strategies in cochlear implantation," Am. J. Otolaryngol. **16**, 669–681.
Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., and Choe, A. (**2008**). "A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/," J. Acoust. Soc. Am. **123**, 4466–4481.

Identification of frequency-shifted vowels

Peter F. Assmann[a)]
School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688

Terrance M. Nearey[b)]
Department of Linguistics, University of Alberta, Edmonton, AB T6G 2E7, Canada

(Received 16 January 2008; revised 26 June 2008; accepted 1 August 2008)

Within certain limits, speech intelligibility is preserved with upward or downward scaling of the spectral envelope. To study these limits and assess their interaction with fundamental frequency (F0), vowels in /hVd/ syllables were processed using the STRAIGHT vocoder and presented to listeners for identification. Identification accuracy showed a gradual decline when the spectral envelope was scaled up or down in vowels spoken by men, women, and children. Upward spectral envelope shifts led to poorer identification of children's vowels compared to adults, while downward shifts had a greater impact on men's vowels compared to women and children. Coordinated shifts (F0 and spectral envelope shifted in the same direction) generally produced higher accuracy than conditions with F0 and spectral envelope shifted in opposite directions. Vowel identification was poorest in conditions with very high F0, consistent with suggestions from the literature that sparse sampling of the spectral envelope may be a factor in vowel identification. However, the gradual decline in accuracy as a function of both upward and downward spectral envelope shifts and the interaction between spectral envelope shifts and F0 suggests the additional operation of perceptual mechanisms sensitive to the statistical covariation of F0 and formant frequencies in natural speech. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2980456]

PACS number(s): 43.71.An, 43.71.Es, 43.70.Mn, 43.71.Bp [PEI] Pages: 3203–3212

I. INTRODUCTION

The anatomical structures of speech vary in size as a function of talker age and sex, and as a result listeners are exposed to a range of variation in average fundamental frequency (F0) and formant frequencies when they hear the same speech sound produced by different talkers. As children develop into adults, increases in the length of the vocal tract result in a progressive lowering of the formant frequencies, and increases in the length and mass of the vocal folds lead to lower average F0 (Vorperian and Kent, 2007). Vocoder studies have indicated that speech remains intelligible when the formant frequencies are scaled up or down (i.e., shifted in log frequency) across a fairly wide range, but intelligibility breaks down with more extreme shifts. Limits on the spectral transposability of speech may impose important constraints on models of talker normalization (e.g., Joos, 1948; Nearey, 1989; Traunmüller, 1994; Irino and Patterson, 2002; Johnson 2005).

One of the earliest studies of the effects of frequency shifts on speech intelligibility was conducted by Chiba and Kajiyama (1941). They varied the playback speed of a phonograph and found that intelligibility was preserved over a range of about 80%–150%, and that slowing down a male voice or speeding up the voices of women or children led to lower intelligibility. However, a limitation of Chiba and Kajiyama's technique is that it applies the same proportional shift to the formant frequencies and F0, which results in a substantial departure from the covariation of F0 and formant frequencies in natural voices. In natural speech, the change from male to female speech involves an upward shift of about 15%–20% in the formant frequencies, accompanied by an 80%–100% increase in F0 (e.g., Peterson and Barney, 1952).

Fu and Shannon (1999) used a channel vocoder to examine the identification of frequency-shifted vowels and found that accuracy followed an inverted U-shaped function, with a fairly steep decline for spectral shift factors smaller than 0.6 or greater than 1.5. Their results indicated that vowels spoken by men were somewhat more resistant to upward shifts in the spectral envelope than vowels of women and children, while children's vowels were more resistant to downward shifts compared to those of adults. Fu and Shannon used noise excitation in a channel vocoder with reduced spectral resolution to simulate the effects of cochlear implant processing and did not investigate the contribution of F0.

A number of studies have shown that vowel quality is affected by a change in F0 (e.g., Miller, 1953; Traunmüller, 1981; Ryalls and Lieberman, 1982; Johnson, 1990; Nearey, 1989; Diehl *et al.*, 1996). For example, Slawson (1968) reported that the quality of a synthetic vowel was altered when F0 was increased by an octave, but these changes were minimized when a simultaneous 10% increase in the formants F1 and F2 was applied. Other studies have shown that vowels synthesized with an inappropriate F0 (relative to formant ranges) may be less accurately identified (Lehiste and Meltzer, 1973; Ryalls and Lieberman, 1982). However, when F0 is high, a special problem arises because the spectral envelope is sparsely sampled at the frequencies of the harmonics,

[a)]Electronic mail: assmann@utdallas.edu
[b)]Electronic mail: t.nearey@ualberta.ca

TABLE I. Geometric mean formant frequencies (F1, F2, and F3) and F0 sampled at 20% of the vowel duration for the 11 vowels from each of the talkers used in the experiments (as indicated in the bottom row).

	Males			Females		Children	
	Talker 1	Talker 2	Talker 3	Talker 1	Talker 2	Talker 1	Talker 2
Geometric mean F0	121	90	115	218	222	239	259
Geometric mean (F1,F2,F3)	1233	1213	1244	1491	1525	1652	1670
Experiments	1,2,3	1,2,3	1,2	2	2	2	2

with the result that the formant peaks may be poorly resolved (Ryalls and Lieberman, 1982; de Cheveigné and Kawahara, 1999; Katz and Assmann, 2001).

Several models assign a direct role for F0 in vowel perception (e.g., Syrdal and Gopal, 1986; Miller, 1989) while others propose that it plays an indirect role, for example by providing information about the characteristics of the talker and the expected range of the formant frequencies (e.g., Nearey, 1989; Traunmüller, 1994; Johnson, 2005; Nusbaum and Magnuson, 1997). In whispered speech, and in experiments such as those reported by Fu and Shannon (1999) with noise-excited speech, it is clear that vowel identification can be maintained even when F0 is not present. However, there is a moderate correlation between F0 and formant frequencies in vowels associated with size differences in the larynx and vocal tract across talkers (Peterson and Barney, 1952). If listeners are sensitive to this covariation, then it might be predicted that identification accuracy would be higher when the relationship between F0 and formant frequencies for a given speaker approximates the observed covariation (e.g., high F0 accompanied by low formant frequencies) and lower when these properties are mismatched (e.g., low F0 combined with high formants). The aim of the present study was to investigate listeners' sensitivity to the covariation of F0 and formant frequencies in the identification of frequency-shifted vowels. The experiments reported here investigated the separate and combined effects of shifts in F0 and spectral envelope on the identification of vowels in /hVd/ syllables. We used a high-quality vocoder, STRAIGHT (Kawahara, 1997; Kawahara *et al.*, 1999) to shift the F0 and spectral envelope of the vowels (resulting in a multiplicative scaling of the formant frequencies) and presented the frequency-scaled versions to listeners for identification. Experiment 1 investigated the effects of upward shifts in male voices. Experiment 2 examined both upward and downward shifts in male, female, and child voices. Experiment 3 used a wider range of scale factors with male voices. In experiments 2 and 3 we introduce a statistical framework that accurately predicts the key aspects of the results. Overall, the results indicate that identification accuracy is affected by the natural ranges of F0 and formant frequencies in human voices and by the relationship between the two.

II. EXPERIMENT 1

A. Method

1. Stimuli

Frequency-scaled vowels were constructed by processing a set of 11 vowels in /hVd/ words spoken by three adult males, drawn from a larger sample of vowels recorded from 10 men, 10 women, and 30 children, ages 3, 5, and 7 years from the North Texas region (Assmann and Katz, 2000; Katz and Assmann, 2001). The left portion of Table I shows the geometric mean F0 and formant frequencies for the three male talkers. The 11 vowels were /i/ (*heed*), /ɪ/ (*hid*), /e/ (*hayed*), /ɛ/ (*head*), /æ/ (*had*), /ʌ/ (*hud*), /ɝ/ (*herd*), /ɑ/ (*hod*), /o/ (*hoed*), /ʊ/ (*hood*), /u/ (*who'd*). The vowels were extracted from the /hVd/ syllables, starting at the first pitch period of the voiced portion of the syllable, and terminating at the end of the last pitch period before the stop closure associated with the final /d/. From these original signals, 15 synthesized versions of each vowel were produced using the STRAIGHT vocoder. These included five spectral envelope scale factors (1.0, 1.25, 1.5, 1.75, and 2.0) combined with three F0 scale factors (1.0, 2.0, and 4.0). When the spectral envelope and F0 scale factors were both set to 1.0, the synthesized versions were virtually indistinguishable from the original recordings (see Sec. V B). Scale factors were selected on the basis of analyses of vowel formant frequencies and F0 combined with pilot listening. The largest frequency shifts produced sounds like those spoken by small children; intermediate shifts were heard as female voices; small shifts were heard as adult males.

2. Listeners

Eleven young adult listeners with normal hearing were recruited from the undergraduate Psychology subject pool at the University of Texas at Dallas and received partial course credit for their participation. All of the listeners were native speakers of American English with no pronounced regional accent; 10 of the 11 were long-term residents of the North Texas region. They were required to complete a dialect questionnaire and a hearing screen (pure tone thresholds better than 25 dB HL at octave frequencies between 0.25 and 8.0 kHz).

3. Procedure

Stimuli were presented at a sample rate of 48 kHz (Tucker-Davis Technologies FT6-2, PA4, HB6) to the right ear via TDH-39 headphones in a double-walled sound booth. Vowels were presented at an average level of 67 dB SPL (A-weighting) with a range between 64 and 70 dB. Listeners identified the vowels using an 11-button response window displayed on the computer screen. Each panel was labeled with the orthographic representation of the /hVd/ syllable and a keyword containing the same vowel: /i/ ("heed," "weed"); /ɪ/ ("hid," "did"); /e/ ("hayed," "made"); /ɛ/

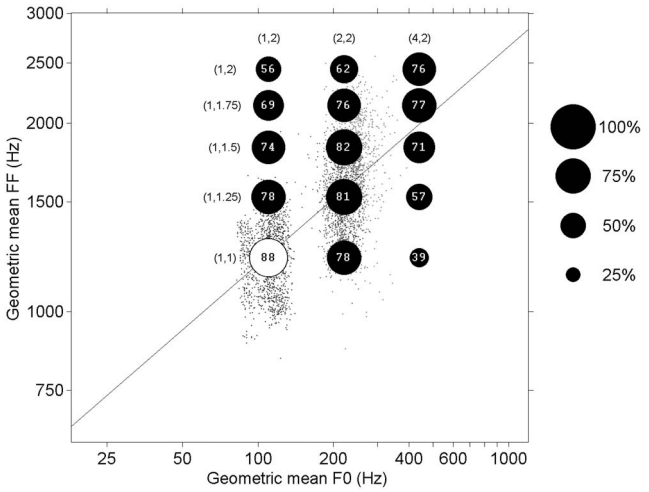


FIG. 1. Mean vowel identification accuracy as a function of F0 and spectral envelope scaling in experiment 1. Circle size is proportional to identification accuracy; circle position indicates the scaled location of the original vowels in the space defined by geometric mean F0 and the geometric mean of the formants F1, F2, and F3. Scale factors are indicated in parentheses (F0 scale factor on the left, spectral envelope on the right). The small dots show acoustic measurements of individual vowels, along with the regression line for predicting log-averaged formant frequencies from log-averaged F0.

("head," "bed"); /æ/ ("had," "dad"); /ʌ/ ("hud," "flood"); /ɑ/ ("hod," "cod"); /o/ ("hoed," "toad"); /ʊ/ ("hood," "wood"); /u/ ("who'd," "rude"); and /ɝ/ ("herd," "bird").

Prior to the experiment, listeners completed practice sessions until they reached a vowel identification score of 86% (19/22) or higher on a set of /hVd/ syllables spoken by a talker who was not included in the test trials. In the main experiment listeners heard 495 stimuli (11 vowels × 3 talkers × 5 spectral envelope shifts × 3 F0 shifts). All conditions, vowels, and talkers were randomly interspersed.

B. Results and discussion

Figure 1 plots the results in the space defined by acoustic measurements of the stimuli. The abscissa shows the geometric mean F0 of the stimuli, while the ordinate shows the geometric mean of the frequencies of the lowest three formants (labeled as FF). F0 measurements were log-transformed and averaged across all voiced frames, 11 vowels and 3 talkers. Measurements of F1, F2, and F3 were taken at the 20% time point in the voiced portion of the vowel, log-transformed and averaged across formants, vowels, and talkers. The open circle shows the average location of the baseline stimuli (F0 and spectral envelope scale factors of 1.0) in this space. The circle centers for the other conditions were scaled up by the indicated F0 and FF scale factors. The size of the circle is proportional to mean identification accuracy, and the number inside the circle indicates the percent correct score. The dots show measurements for individual vowels from 50 speakers (20 adults and 30 children) from the north Texas region (Assmann and Katz, 2000). The dashed line shows the regression line predicting log-averaged formant frequencies from log-averaged F0 for these individual points.

The left column of circles in Fig. 1 shows that identification accuracy declined progressively as the spectral envelope scale factor was increased. Similarly, identification accuracy declined when F0 was increased and the spectral envelope scale factor was unchanged (bottom row). However, performance remained high when F0 and spectral envelope scale factors were both increased (upper right, near the regression line).

Identification scores were converted to rational arcsine units (RAUs) (Studebaker, 1985). Analysis of variance on RAU-transformed scores indicated a significant effect of F0, $F(2,20)=45.33$; $p<0.01$, and spectral envelope scale factor, $F(4,40)=8.22$; $p<0.01$, and a significant two-way interaction of F0 and spectral envelope scale factors, $F(8,80)=40.80$; $p<0.01$. Correlated means Šidak-corrected t-tests were conducted using pairwise within-subject comparisons (Myers and Well, 1991, p. 250) of each of the 14 shifted conditions with the unshifted baseline. These tests indicated the following: (1) for unshifted F0, all upward shifts in spectral envelope led to significantly lower accuracy; (2) for F0 scaled by a factor of 2.0, the two largest spectral envelope shifts led to significantly lower accuracy; and (3) for F0 scaled by 4.0, all spectral envelope shifts except the largest (2.0) resulted in lower accuracy. Planned comparisons indicated that performance was better when the largest upward shift in F0 was combined with the largest upward shift in spectral envelope, compared to the case where (1) the largest shift in F0 was coupled with an unshifted spectral envelope, $t(10)=7.82$; $p<0.01$; and (2) the largest shift in spectral envelope was combined with unshifted F0, $t(10)=14.24$; $p<0.01$.

Previous studies have reported lowered accuracy for upward shifts in spectral envelope (Fu and Shannon, 1999). One explanation for this decline is that spectral envelope shifts that produce signals with formant frequencies far outside the natural human range may sound unfamiliar or less speech-like, making them harder to interpret. Compatible with this idea, identification accuracy is also lower when F0 is increased well above the natural range for male voices while holding the formant pattern constant within the male range. When the formants and F0 are both shifted up together, in a manner similar to that found in children's voices, performance is better than when either dimension is scaled separately. Performance is poorest in conditions where the formant pattern and F0 violate expected patterns in natural speech (i.e., away from the regression line in Fig. 1).

A second explanation for the interaction of F0 and spectral envelope is that performance may drop because the representation of the first formant is poor when F0 is high. For example, it has been suggested that vowel quality deteriorates when F0 is higher than the frequency of the first formant (Smith *et al.*, 2005). Thus, the finding that performance improves as spectral envelope and F0 are raised together could be attributed to the enhanced representation of the first formant.[1]

If the second account is correct, then the advantage for matched conditions should appear only with upward shifts, not with downward shifts of F0. Experiment 2 tested this prediction by applying both upward and downward shifts in F0 and spectral envelope. To extend the generality of these effects, the frequency-shifted stimuli were produced from re-

corded vowels spoken by two men, two women and two children. Fu and Shannon (1999) reported that vowel identification was more adversely affected by upward shifts in spectral envelope for children compared to adults, and that male vowels were less accurately identified than female vowels for downward shifts. However, their stimuli were constructed using a noise-excited channel vocoder, which effectively removes F0 (i.e., the F0-related periodicity and harmonic structure) from the signal. Experiment 2 sought to replicate and extend Fu and Shannon's findings using voiced vowels and included F0 manipulations to measure the interaction between spectral envelope and F0 shifts in adult male, female, and children's voices.

III. EXPERIMENT 2

A. Method

1. Stimuli

The stimuli were 11 vowels in /hVd/ words spoken by two adult males, two adult females, and two children aged 7 years from our North Texas vowel database. The male talkers were a subset of those used in experiment 1. Table I shows the geometric mean F0 and formant frequencies for the six talkers. The vowel portions were extracted as described in experiment 1 and frequency-shifted versions were constructed using the STRAIGHT vocoder. Five spectral envelope scale factors (0.6, 0.8, 1.0, 1.5, and 2.0) were combined with three F0 scale factors (0.5, 1.0, and 4.0).

2. Listeners

The listeners were 14 monolingual speakers of American English, long-term residents of North Texas with no reported history of speech, hearing, or language disorders. None had participated in experiment 1.

3. Procedure

The stimuli were presented diotically over Sennheiser HD-540 earphones following a procedure similar to that of experiment 1. Listeners labeled the vowels using an 11-category response box on the computer screen. Prior to the experiment, listeners completed practice sessions on a separate vowel set until they reached a mean identification score of 85% correct. Listeners identified 990 stimuli (11 vowels, 6 talkers, 5 spectral envelope shifts, and 3 F0 shifts). All stimulus conditions were randomly interspersed.

B. Results and discussion

Figure 2 indicates reduced identification accuracy in several conditions for all three talker groups when the spectral envelope was scaled upwards by a factor of 2.0 (top row in each panel) or downwards by a factor of 0.6 (bottom row in each panel) compared to the unscaled baseline condition (unfilled circles). Consistent with Fu and Shannon (1999), children's vowels were less adversely affected by downward shifts than those of adult males, while adult male vowels were more resistant to upward shifts compared to children. Since the formant frequencies of children are higher than those of adults, this result suggests that there may be a limit

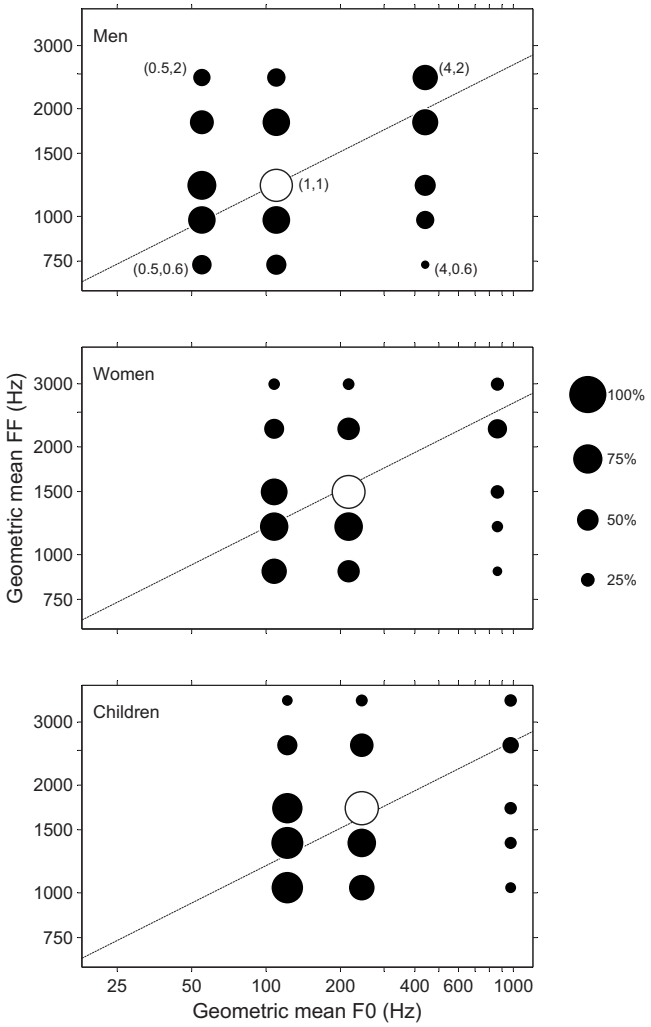


FIG. 2. Mean vowel identification accuracy as a function of F0 and spectral envelope scaling for men's (upper panel), women's (middle panel), and children's vowels (bottom panel) in experiment 2. Circle size is proportional to identification accuracy. Circle position indicates the scaled location of the original vowels (averaged across 11 vowels and two talkers) in the space defined by geometric mean F0 (abscissa) and the geometric mean of the formants F1, F2, and F3 (ordinate). Scale factors are indicated in parentheses (F0 on the left, spectral envelope on the right); the regression line in each panel is the same as in Fig. 1.

on the extent of spectral envelope shifts that preserve vowel quality. It is noteworthy, however, that the drop in intelligibility is gradual and in most cases remains well above the chance level of 1/11 (approximately 9%).

Vowels in the baseline condition were well identified, as indicated by the unfilled circles. In general, performance was higher in conditions where spectral envelope and F0 shifts produced stimuli near the regression line derived from the measurements of natural vowels, compared to points more distant from the regression line. Stimuli that mirror the natural covariation of F0 and formant frequencies (i.e., lie close to the regression line) were generally better identified than those that do not follow this pattern.

Identification accuracy scores were RAU-transformed and subjected to a three-way repeated measures analysis of variance, covering the factors spectral envelope shift, F0 shift, and talker group. All main effects and interactions were

significant. Of key interest was the three-way interaction of spectral envelope shift, F0 shift, and talker group [$F(16,208)=4.82$; $p<0.01$]. *A priori* tests revealed that increasing the spectral envelope scale factor to 2.0 (without concomitant changes in F0) led to significantly lower accuracy for vowels spoken by men, $t(13)=9.89$; $p<0.01$ confirming the results of experiment 1. The same pattern was seen for women's vowels, $t(13)=13.48$; $p<0.01$ and children's vowels $t(13)=23.91$; $p<0.01$. As in experiment 1, for men's vowels (but not those of women and children) the combination of shifting the spectral envelope up by a factor of 2.0 and F0 by a factor of 4.0 produced significantly higher identification accuracy than shifting the spectral envelope alone, $t(13)=9.18$; $p<0.01$. Lowering the spectral envelope by applying a scale factor of 0.6 (without changing F0) also produced lower accuracy for all three talker groups [men: $t(13)=7.25$; $p<0.01$; women: $t(13)=7.88$; $p<0.01$; children: $t(13)=7.72$; $p<0.01$]. However, lowering the spectral envelope (scale factor of 0.6) in combination with a lowered F0 (scale factor of 0.5) led to significantly improved scores compared to the case where only the spectral envelope was lowered for women's vowels, $t(13)=6.55$; $p<0.01$, and children's vowels, $t(13)=4.84$; $p<0.01$.

Overall, the results can be summarized as follows: First, raising or lowering the spectral envelope generally led to lower identification accuracy. Upward shifts had more harmful consequences for children's vowels while downward shifts had a greater negative impact on vowels spoken by men. The reduction in accuracy was ameliorated in some cases by concomitant changes in F0. Second, increasing F0 by a factor of 4.0 generally led to lower identification accuracy. In several conditions this drop was counteracted by increasing the spectral envelope scale factor, most notably for men's vowels.

There are two possible explanations for the compensatory interaction of F0 and spectral envelope. First, an increase in F0 results in a spectrum with fewer harmonics, which may lead to an impoverished representation of the spectral envelope. When F0 is high and the spectral envelope is shifted down, the frequency of the first formant may be below F0. Upward scaling of the spectral envelope can improve the representation of F1 by providing closer alignment of harmonics with the first formant peak. Consistent with this idea, an increase in spectral envelope scale factor from 0.6 to 1.5 led to a progressive improvement from near-chance levels (9%) to about 40%–50% correct for vowels spoken by women and children with an F0 shift factor of 4.0.

A second reason for the interaction is that coordinated shifts in formant pattern and F0 produce a better match to the expected pattern in natural voices. Voices with a low formant pattern combined with high F0 (or *vice versa*) are perceived as unnatural (Assmann *et al.*, 2006; Assmann and Nearey, 2007a). Listeners may incorporate prior expectations about F0 and formant ranges to determine the phonetic properties of frequency-shifted vowels (Nearey and Assmann, 2007).

Summary model

To obtain a rough quantitative estimate of the importance of mismatch between F0 and formant pattern in the data represented in Fig. 2, we constructed a multiple regression model on the pooled responses from all subjects as represented by the circle sizes. The model incorporated the following independent variables (related to terms in the analysis of variance reported above): (1) linear and quadratic effects of the average log F0 (the log of the *x*-axis of the right panels in Fig. 2) to allow for the possibility of a general increase or decrease in performance with increasing F0, as well as the possibility that extremely large or small values of F0 might perform differently than average values; (2) linear and quadratic trends for the average log formant frequencies (the log of the *y*-axis of the right panels in Fig. 2) and (3) distance (of the observed geometric mean formant frequency, in log units) from the regression line (related to the interaction of F0 and formant frequencies). In the full regression model, all factors were nominally significant, $F(1,39)=21.79$; $p<0.0001$.

The full model with six degrees of freedom in the predictor accounted for approximately 93% of the total variance. The relative contributions of F0 and formant frequencies and their interaction (distance to regression line) can be assessed by dropping the corresponding terms from the model. From this perspective, the distance D to the regression line makes the largest contribution, followed closely by F0, with formant frequencies showing considerably smaller effects. Dropping D from the full model decreases R^2 by about 26 percentage points. Dropping the linear and quadratic F0 terms decreases R^2 by about 25 percentage points. Dropping the linear and quadratic formant frequency terms decreases R^2 by only about 7 percentage points to 79%. The modeling results confirm that the interaction of F0 and formant frequencies (as represented by D) makes a significant independent contribution to identification accuracy.

Experiment 3 explored this contribution further by investigating a rectangular grid of spectral envelope and F0 shift factors extending beyond the range found in the voices of adults and children. If the degree of mismatch between the formant pattern and F0 is an important factor in vowel identification, it is predicted that identification accuracy will be higher for matched shifts (e.g., high F0 and high formant frequencies) than for mismatched shifts (e.g., high F0 and low formants), even when both parameters lie outside the values observed in natural speech.

IV. EXPERIMENT 3

A. Method

1. Stimuli

Two of the three adult males selected in experiment 1 provided the seed voices. The 11 vowels in /hVd/ words were processed using STRAIGHT to generate 36 synthesized versions. These included six F0 scale factors (0.25, 0.5, 1.0, 2.0, 4.0, and 8.0) combined with six spectral envelope scale factors (0.61, 0.78, 1.00, 1.28, 1.64, and 2.10). Small adjustments were applied to the scale factors so that the two selected talkers' vowels in the unshifted condition were centered on the geometric means for male talkers in the north Texas vowel database; this point corresponds to the center of

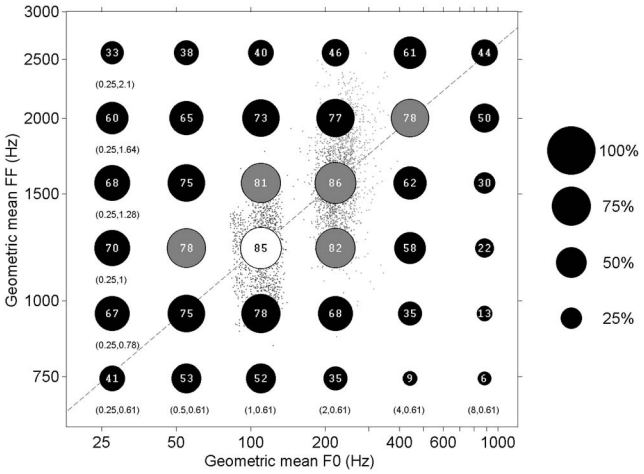


FIG. 3. Mean vowel identification accuracy as a function of F0 and spectral envelope scaling in experiment 3. Circle size is proportional to identification accuracy; circle position indicates the scaled location of the original vowels (averaged across 11 vowels and two talkers) in the space defined by geometric mean F0 (abscissa) and the geometric mean of the formants F1, F2, and F3 (ordinate). Scale factors are indicated in parentheses (F0 on the left, spectral envelope on the right). Measurements of natural vowels (small dots) and regression line are the same as in Fig. 1.

the white circle in Fig. 3. The remaining scale factors were selected to provide equidistant logarithmic spacing across the natural human range and beyond.

2. Listeners

Twenty students with normal hearing were recruited from the undergraduate Psychology subject pool at the University of Texas at Dallas, and received partial course credit for their participation. All were native speakers of American English with no pronounced regional accent. None had previously taken part in experiments using synthesized vowels or frequency-shifted speech.

3. Procedure

Stimuli were presented diotically via Sennheiser HD-540 earphones following the procedure described in experiment 1. Listeners completed practice sessions until they reached a score of 86% (19/22) or better on a set of /hVd/ words spoken by a different talker. In the main experiment they heard 792 stimuli (11 vowels $\times$ 2 talkers $\times$ 6 spectral envelope shifts $\times$ 6 F0 shifts), with all conditions randomly interspersed.

B. Results and discussion

The large circles in Fig. 3 show identification accuracy for the frequency-scaled vowels as a function of the F0 and spectral envelope scale factors, averaged across the 20 listeners, 2 talkers, and 11 vowels. It can be seen that identification accuracy was generally higher for scale factors comparable to those found in measurements of natural vowels, near the regression line. Performance declined progressively with increasing distance from the regression line, with the steepest drop for vowels with raised F0 and lowered spectral envelope scale factor (bottom right corner of Fig. 3).

An analysis of variance on RAU-transformed identification accuracy scores indicated significant main effects of spectral envelope scale factor [$F(5,95)=167.44$; $p<0.01$] and F0 scale factor [$F(5,95)=295.26$; $p<0.01$] as well as a significant interaction, $F(25,475)=57.31$; $p<0.01$. Šidak corrected tests of all treatments against controls (family size=35) were conducted on pairwise within-subject differences. The white circle in Fig. 3 shows the unshifted baseline condition, while the gray circles show all of the conditions that did not differ significantly from the baseline ($p>0.05$); all remaining conditions are shown in black. From this comparison it can be seen that identification accuracy is generally higher for matched combinations, in which F0 and spectral envelope scale factors are both raised or both lowered, compared to conditions where one is raised and the other is lowered. The improved accuracy for matched combinations is qualitatively consistent with the natural co-variation of F0 and spectral envelope scale factors across talkers. However, the results also indicate that accurate vowel identification can be obtained with scale factors that lie outside the range observed in the vowel database (see also Smith *et al.*, 2005).

Several more focused, and hence, more powerful, tests of contrasts suggested by analyses in experiments 1 and 2 were also conducted. With an F0 scale factor of 1.0, Šidak tests (family size=5) indicated a significant drop in accuracy for upward shifts in spectral envelope with scale factors of 1.64 and 2.10 and downward scale factors of 0.78 or 0.61 [all $F(1,19)>22.92$, Šidak corrected $p<0.007$]. Similarly, increasing F0 by a factor of 4.0 or 8.0 or decreasing it by factors of 0.5 or 0.25 produced a significant drop when the spectral envelope scale factor was held at 1.0 [all $F(1,19)>13.25$, Šidak corrected $p<0.009$]. Finally, considering the upward scale factors closest to the largest used in experiments 1 and 2, we found a significant improvement when F0 and spectral envelope were both shifted up (scale factors of 4.0 and 1.64, respectively) compared to conditions where only the envelope was shifted [$t(19)=8.845$, $p<0.00001$]. A non-significant improvement in the expected direction was found comparing the coordinated shift to the case where only F0 was shifted [$t(19)=1.52$, $p>0.07$, one-tailed].

Summary model

To gauge the relative importance of the F0—formant mismatch for identification accuracy we applied the multiple regression model developed in experiment 2 to the pooled responses from all subjects as represented by the circle sizes. This model includes as independent variables linear and quadratic effects of the average log F0, linear and quadratic trends for formant frequencies, and distance from the regression line (related to the interaction of F0 and formant frequencies). The full model with six degrees of freedom in the predictor accounted for 92% of the variance. In the full regression model, all factors were nominally significant[2] [$F(1,30)\geq 16.02$; $p<0.0001$]. The full model accounted for approximately 92% of the total variance. The relative contributions of the main effect (F0, formant frequencies) factors and the interaction factor (distance to regression line) can be assessed by dropping the related terms from the model. From this perspective, dropping the linear and quadratic F0 terms

has the largest effect, decreasing R^2 by about 27 percentage points. Dropping distance to the regression line D from the full model decreases R^2 by about 19 percentage points. Dropping the linear and quadratic formant frequencies terms also decreases R^2 by about 19 percentage points.

The balanced structure of the stimuli in experiment 3 makes it possible to decompose the regression into three components that lend themselves to tentative interpretation. The four contour plots in Fig. 4 show the changes in the predicted identification accuracy due to each of these components along with the full prediction surface.

Panel (a) shows predicted identification accuracy (in RAU units) for linear F0 and formant frequency effects. This component may be viewed tentatively as roughly encoding a spectral density factor. Relative to the unscaled condition, stimuli in the upper left quadrant exhibit compression of the harmonic spacing, combined with stretching the spectral envelope, thereby increasing the number of harmonics per formant. Stimuli in the lower right quadrant show compression of the spectral envelope and an increase in harmonic spacing, resulting in fewer harmonics per formant. Panel (b) shows the effects of quadratic trends in F0 and formant frequency.[3] This component reflects the trend for scale values near 1.0 to be more accurately identified than more extreme F0 and formant frequency values. Panel (c) shows the effects of distance from the regression line and reflects the tendency for lower accuracy with increased distance from the regression line. Panel (d) shows the predicted RAU scores for the combined model, which is the sum of the first three panels and a constant term. The combined effect of the putative spectral density effect of panel (a) and distance from the regression line in panel (c) are responsible for the asymmetry below the regression line shown in panel (d).[4]

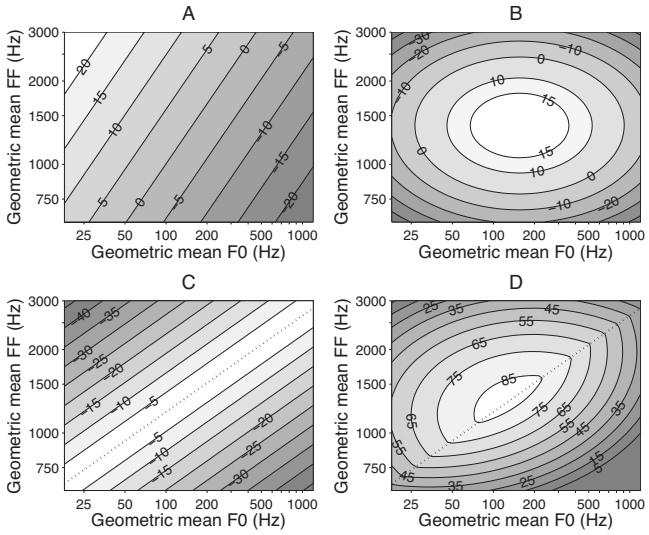


FIG. 4. Components of descriptive model for experiment 3. The z-axis represents changes in RAU scores.

V. GENERAL DISCUSSION

A. Frequency shifts and vowel identification

In each of the experiments, identification accuracy remained at fairly high levels across a wide range of spectral envelope and F0 scale factors. However, performance showed a gradual and steady decline as the spectral envelope scale factor was increased or decreased relative to the unshifted original. Experiment 2 confirmed previous reports (e.g., Fu and Shannon, 1999) that men's vowels are more susceptible to downward shifts compared to vowels spoken by women and children, while children's vowels showed a greater decline with upward shifts. This suggests that the performance decline is associated with the absolute ranges of the formant frequencies (or related features of the spectral envelope) across age/sex classes. Since men's vowels have lower formant ranges than those of women and children, a plausible interpretation is that the performance decline is a reflection of listeners' experience with natural speech and their lack of exposure to such extreme formant patterns. Consistent with this interpretation, studies of frequency-shifted speech have found that intelligibility improves with extended exposure (Rosen *et al.*, 1999; Nogaki *et al.*, 2007).

Smith *et al.* (2005) have suggested that vowels with high F0s may be less accurately recognized for a different reason, namely the poor representation of the first formant. When F0 is increased, the number of harmonics defining the first formant is reduced, and in extreme cases, F0 may actually be higher than the frequency of the first formant. This factor may help to explain the progressive decline in identification accuracy when F0 is increased and the spectral envelope scale factor is decreased, which results in F0 being higher than F1 for some stimuli. However, this does not explain the significant drop in accuracy when the spectral envelope scale factor is raised, or the interaction between spectral envelope and F0 for downward shifts. It is possible that some aspects of the results may be related to a spectral density factor, which could include underspecification of F1 at high F0, as discussed in the interpretation of Fig. 4. However the discrepancy in patterns for the linear F0 and formant frequency effects in experiments 2 and 3 (see note 4) and the absence of a coherent perceptual account of the effects of a sparsely sampled harmonic spectrum make it seem unwise to draw conclusions on this issue without additional evidence.

A pattern that emerges from all three experiments is that vowel identification is more accurate when F0 and spectral envelope scale factors are shifted in the same direction rather than in opposite directions. In some conditions this pattern leads to a slight "recovery" function (i.e., improved performance when F0 and spectral envelope are both scaled up (for adult males) or both scaled down (for children) compared to conditions where only one dimension is scaled. F0 may provide a benefit in these extreme cases because it helps to delineate the properties of the talker's voice and reduces uncertainty about the identity of the vowel in a region where listeners have limited exposure. The finding that the recovery function only holds over a selected range is also consistent with a role of listening experience to natural voices (Nearey and Assmann, 2007).

B. Synthesis fidelity

The question arises whether reduction in identification accuracy in frequency scaling studies of this type is due to

shortcomings of the synthesis procedure rather than perceptual constraints. Given that STRAIGHT was specifically designed and developed for voice manipulations of this type (Kawahara *et al.*, 1999), it seems unlikely that synthesis artifacts are responsible for substantial decrement in accuracy, at least within the range of observed variation in F0 and formant frequencies in adult and child voices. Identification accuracy for unscaled, STRAIGHT-vocoded vowels closely matches accuracy rates for natural vowels spoken by men, women and children (Assmann and Katz, 2005). STRAIGHT provides quite compelling changes between male and female or adult and child voices. When the male vowels in experiments 1–3 were scaled to child ranges in both F0 and formant frequencies, the identification rates were comparable to those previously observed in children's natural vowels (Assmann and Katz, 2005). Similarly, in experiment 2, children's vowels were scaled to adult ranges and were identified quite accurately. Finally, we have confirmed that spectral envelope and F0 scaling using STRAIGHT allows us to recover the scaled F0 and formant frequencies across a range of scale factors. Difficulties arise in formant estimation when F0 is high relative to the formant frequencies. However, such difficulties may not be simply a problem of synthesis fidelity; listeners may face related problems (see Dissard and Darwin, 2001). It may be impossible to isolate the role of synthesis fidelity for scaling outside the natural human range when there is no objective way to assess the performance of the synthesizer that is independent of the perceptual mechanisms under investigation.

C. Models of talker normalization

Irino and Patterson (2002) have proposed that a scale transform is applied to the internal representation of sound at an early stage of auditory analysis prior to categorization. The scale transform, which applies equally to speech and other types of sounds, effectively partitions size information (associated with vocal tract length differences across talkers) from shape information (associated with differences in vowel quality). According to the model, information about the talker is segregated from vowel category information at a stage prior to and independently of vowel recognition and provides a basis for age/sex classification associated with differences in vocal tract length. In support of this hypothesis, Smith *et al.* (2005) demonstrated that listeners can discriminate small changes in talker size and can recognize scaled vowels over a wide range of spectral envelope and F0 scale factors, extending well beyond the range encountered in natural speech.[5] The present results provide further evidence that vowel identification remains at a high level across a fairly wide range of scale factors. However, several aspects of the results are difficult to reconcile with the idea that talker normalization takes place prior to and independently of vowel categorization as proposed by Irino and Patterson (2002).

First, as shown in Fig. 3, vowel identification shows a steady and progressive decline as the spectral envelope scale factor is increased or decreased relative to the natural speech range. Second, identification accuracy interacts with F0: scaled vowels that lie near the regression line in Figs. 1–3 are identified more accurately than those more distant. The finding that vowel identification is affected by the natural ranges of F0 and formant frequencies and their interaction is incompatible with a scale transform that precedes vowel identification.

Smith *et al.* (2005) have argued that the effects of frequency shifts on vowel identification cannot be explained by statistical pattern recognition models that rely on natural relationships between formant frequencies and F0 in natural speech because such models have no basis for extrapolating beyond the range of natural voices and would therefore predict a sharp deterioration in performance for frequency shifts outside this range. However, this is clearly not true in general for statistical vowel recognition models using linear discriminant analysis or related Bayesian classifiers (e.g., Nearey and Assmann, 1986; Gottfried *et al.*, 1993; Zahorian and Jagharghi, 1993; Hillenbrand *et al.*, 2001). Such methods apply in principle to unbounded pattern spaces and automatically generate predictions beyond the ranges of the training data. Nearey and Assmann (2007) have recently demonstrated that such models of vowel identification trained on F0 and formant measurements of natural vowels spoken by men, women, and children can indeed classify frequency-shifted vowels successfully even when the scale factors exceed the natural speech range. Furthermore, the models are able to predict aspects of the interaction between F0 and spectral envelope shifts, including the overall decline with more extreme shifts in spectral envelope and the increased accuracy for matched shifts (e.g., an increase in spectral envelope combined with increase in F0) compared to mismatched shifts (F0 and spectral envelope shifted in opposite directions). Although the models successfully predicted many aspects of listeners' overall correct identification rates, they generally predicted better performance for matched combinations of F0 and spectral envelope scale factors than observed in listeners and worse performance than observed for mismatched combinations. Furthermore, discrepancies were found in the pattern of predicted and observed confusion errors in some cases. That is, even where overall identification rates were relatively well matched, the dominant categories of the predicted errors did not always match listeners' specific error choices. Whether or not refinements of this general class of models can account for the detailed pattern of listeners' responses, there is evidence from a number of sources that the role of F0 may be related to an indirect normalization of the formant space (Johnson, 1990; Nusbaum and Magnuson, 1997; Traunmüller, 1994; Nearey and Assmann, 2007).

An important final question concerns the generalizability of the results for frequency-shifted vowels to the perception of connected speech. We have conducted experiments with frequency-shifted sentences from the H.I.N.T. (Nilsson *et al.*, 1993), using STRAIGHT to apply the same F0 and spectral envelope scale factors as in experiment 3 (Assmann and Nearey, 2007b). Overall percent correct recognition rates for words in sentences were highly correlated with identification rates for vowels in /hVd/ syllables (r=0.92). As in the case of vowels, word recognition scores were higher for matched

shifts than for mismatched shifts, with the lowest performance in conditions where F0 was raised and spectral envelope was shifted down. The parallel results from these two experiments suggest that the interaction between F0 and spectral envelope plays a similar role in vowel identification and word recognition in connected speech.

ACKNOWLEDGMENTS

Thanks to Jack Scott and Tiffani Jantz for assistance in testing the subjects, Hideki Kawahara for providing the STRAIGHT synthesis software, Harald Baayen for advice in statistical matters, Michael Kiefte and two reviewers for helpful comments on the manuscript. Portions of this work were reported at the 141st and 149th Meetings of the Acoustical Society of America, the 7th International Conference on Spoken Language Processing, and the 15th International Congress of Phonetic Sciences. This work was supported in part by grants from the National Science Foundation (0318451) and the Social Sciences and Humanities Research Council of Canada (41020001353).

[1]For example, the number of harmonics between 0 and (F2−F1)/2 Hz is greater when F0 and F1 are both increased, compared to the case where F0 is increased and F1 is lowered or held constant.

[2]After extensive exploration with the data of experiment 2, we decided not to pursue the issue of cross-listener generalization there. However, having decided on the form of the model in experiment 2, we did pursue this issue in the analysis of experiment 3. A series of likelihood ratio tests were undertaken using the lmer2 procedure of the lme4 package in *R*. Repeated measures tests on the same fixed-effects model (and with a random effects model with the same terms conditional on subjects and with an unrestricted covariance structure) confirmed that all reported effects were highly significant by this criterion. In particular dropping distance from the regression line, *D*, from the full model led to a likelihood ratio $\chi^2(1)=53.48$; $p<1.0\cdot e^{-12}$. The numbers reported in the text represent the relative contributions of the fixed effects on the responses pooled across listeners compared to all variation across the conditions for the same pooled values.

[3]Dropping the linear terms associated with panel (a) from the model results in a decrease of 22.3 percentage points in R^2; dropping the quadratic terms of panel (b) results in a reduction of 21.6 percentage points.

[4]Examination of similar plots of components of experiment 2 casts some doubt on the generality of the interpretation as a spectral density factor. Although panels (b) and (c) and the net effects in panel (d) are similar in the two experiments, the sign of the linear formant frequency component in panel (a) is actually reversed in experiment 2. Further research is required to elucidate this issue.

[5]Smith *et al.* (2005) obtained higher levels of identification accuracy for frequency-scaled vowels than those reported here. This discrepancy may be explained in part by the size of the vowel inventory: Smith *et al.* used sustained versions of five English vowels (/i/, /e/, /ɑ/, /o/, and /u/) from a single male talker, while the present study used 11 vowels from men, women and children.

Assmann, P. F., and Katz, W. F. (**2000**). "Time-varying spectral change in the vowels of children and adults," J. Acoust. Soc. Am. **108**, 1856–1866.

Assmann, P. F., and Katz, W. F. (**2005**). "Synthesis fidelity and vowel identification," J. Acoust. Soc. Am. **117**, 886–895.

Assmann, P. F., and Nearey, T. M. (**2007a**). "Relationship between fundamental and formant frequencies in voice preference," J. Acoust. Soc. Am. **122**, EL35–EL43.

Assmann, P. F., and Nearey, T. M. (**2007b**). "Effects of frequency shifts on the identification of vowels and words in sentences," J. Acoust. Soc. Am. **122**, 3064–3065.

Assmann, P. F., Dembling, S., and Nearey, T. M. (**2006**). "Effects of frequency shifts on perceived naturalness and gender information in speech," Proceedings of the Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, September 17–21, **2006**, pp. 889–892.

Chiba, T., and Kajiyama, M. (**1941**). *The vowel: Its Nature and Structure* (Kaiseikan, Tokyo).

de Cheveigné, A., and Kawahara, H. (**1999**). "Missing-data model of vowel identification," J. Acoust. Soc. Am. **105**, 3497–3508.

Diehl, R. L., Lindblom, B., Hoemeke, K. A., and Fahey, R. P. (**1996**). "On explaining certain male-female differences in the phonetic realization of vowel categories," J. Phonetics **24**, 187–208.

Dissard, P., and Darwin, C. J. (**2001**). "Formant-frequency matching between sounds with different bandwidths and on different fundamental frequencies," J. Acoust. Soc. Am. **110**, 409–415.

Fu, Q.-J., and Shannon, R. V. (**1999**). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," J. Acoust. Soc. Am. **105**, 1889–1900.

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

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.

Irino, T., and Patterson, R. D. (**2002**). "Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform," Speech Commun. **36**, 181–203.

Johnson, K. (**1990**). "The role of perceived speaker identity in F0 normalization of vowels," J. Acoust. Soc. Am. **88**, 642–654.

Johnson, K. (**2005**). "Speaker normalization in speech perception," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. Remez (Blackwell, Oxford), pp. 363–389.

Joos, M. A. (**1948**). "Acoustic phonetics," Language **24**, 1–136.

Kawahara, H. (**1997**). "Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited," Proceedings of the ICASSP, pp. 1303–1306.

Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction," Speech Commun. **27**, 187–207.

Katz, W. F., and Assmann, P. F. (**2001**). "Identification of children's and adults' vowels: Intrinsic fundamental frequency, fundamental frequency dynamics, and presence of voicing," J. Phonetics **29**, 23–51.

Lehiste, I., and Meltzer, D. (**1973**). "Vowel and speaker identification in natural and synthetic speech," Lang Speech **16**, 356–364.

Miller, R. L. (**1953**). "Auditory tests with synthetic vowels," J. Acoust. Soc. Am. **18**, 114–121.

Miller, J. D. (**1989**). "Auditory-perceptual interpretation of the vowel," J. Acoust. Soc. Am. **85**, 2114–2134.

Myers, J. L., and Well, A. (**1991**). *Research Design and Statistical Analysis* (Harper Collins, New York).

Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

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., and Assmann, P. F. (**2007**). "Probabilistic 'sliding template' models for indirect vowel normalization," in *Experimental Approaches to Phonology*, edited by M. J. Solé, P. S. Beddor, and M. Ohala (Oxford University Press, Oxford), pp. 246–269.

Nilsson, M., Soli, S. D., 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.

Nogaki, G., Fu, Q. J., and Galvin, J. J. III (**2007**). "Effect of training rate on recognition of spectrally shifted speech," Ear Hear. **28**, 132–140.

Nusbaum, H. C., and Magnuson, J. S. (**1997**). "Talker normalization: Phonetic constancy as a cognitive process," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego), pp. 109–132.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.

Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants," J. Acoust. Soc. Am. **106**, 3629–3636.

Ryalls, J. H., and Lieberman, P. (**1982**). "Fundamental frequency and vowel perception," J. Acoust. Soc. Am. **72**, 1631–1634.

Slawson, A. W. (**1968**). "Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency," J. Acoust. Soc. Am. **43**, 87–101.

Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (**2005**). "The processing and perception of size information in speech sounds," J. Acoust. Soc. Am. **117**, 305–318.
Studebaker, G. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.
Traunmüller, H. (**1981**). "Perceptual dimension of openness in vowels," J. Acoust. Soc. Am. **69**, 1465–1475.
Traunmüller, H. (**1994**). "Conventional, biological, and environmental factors in speech communication: A modulation theory," Phonetica **51**, 170–183.
Vorperian, H., and Kent, R. D. (**2007**). "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," J. Speech Lang. Hear. Res. **50**, 1510–1545.
Zahorian, S., and Jagharghi, A. (**1993**). "Spectral-shape features versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**, 1966–1982.

A model for multitalker speech perception

Soundararajan Srinivasan[a)]
Biomedical Engineering Department, The Ohio State University, Columbus, Ohio 43210

DeLiang Wang[b)]
Department of Computer Science and Engineering and Center for Cognitive Science, The Ohio State University, Columbus, Ohio 43210

(Received 5 November 2007; revised 6 August 2008; accepted 18 August 2008)

A listener's ability to understand a target speaker in the presence of one or more simultaneous competing speakers is subject to two types of masking: energetic and informational. Energetic masking takes place when target and interfering signals overlap in time and frequency resulting in portions of target becoming inaudible. Informational masking occurs when the listener is unable to distinguish target and interference, while both are audible. A computational model of multitalker speech perception is presented to account for both types of masking. Human perception in the presence of energetic masking is modeled using a speech recognizer that treats the masked time-frequency units of target as missing data. The effects of informational masking are modeled as errors in target segregation by a speech separation system. On a systematic evaluation, the performance of the proposed model is in broad agreement with the results of a recent perceptual study. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2982413]

PACS number(s): 43.71.An, 43.72.Dv, 43.72.Ne [DOS] Pages: 3213–3224

I. INTRODUCTION

In everyday listening conditions, the acoustic input reaching our ears is often a mixture of multiple sound sources. In such situations, the human ability to perceive a target source is susceptible to the effects of masking, which is defined as the increase in the audibility threshold of the target (Mayer, 1876). In particular, our ability to attend to and understand a target speaker in the presence of other competing speakers is affected by energetic and informational masking (Tanner, 1958; Brungart *et al.*, 2001). Energetic masking refers to the phenomenon in which a stronger signal masks a weaker one within a critical band (Fletcher, 1940). Recently, the term informational masking has been used to refer to the perceptual degradation caused by the listener's inability to segregate a target signal from interference (Carhart *et al.*, 1969; Pollack, 1975) (for a review, see Watson, 2005). Informational masking, therefore, depends on the similarity of segregation cues such as voice characteristics (Brungart *et al.*, 2001) and spatial locations (Freyman *et al.*, 1999) associated with individual signals. In this paper, we propose a model for recognizing target speech in the presence of both energetic and informational masking under monaural conditions.

Spectrotemporal overlap between target and interference is a prime cause of energetic masking. Portions of a target signal subject to energetic masking become inaudible at the periphery of the auditory system and are unavailable for subsequent processing. A missing-data speech recognizer (Cooke *et al.*, 2001) is therefore used to model speech perception under energetic masking conditions. When target speech is contaminated by additive interference, some time-frequency (T-F) regions are dominated by target energy while some of the rest are dominated by interference. The missing-data method treats the former T-F units as reliable and the latter T-F regions as missing or unreliable during recognition.

The missing-data recognizer requires a T-F mask (typically binary) that provides information about which T-F regions of the mixture signal are reliable and unreliable. The task of generating such a mask is akin to the task of segregating the target from the mixture. The process by which the human auditory system is able to organize the acoustic input into components that correspond to individual sources in the input is known as auditory scene analysis (ASA) (Bregman, 1990). Therefore, informational masking is closely related to ASA. Hence, we adapt a monaural computational ASA (CASA) system to estimate a binary mask that selects those T-F regions of the mixture where the target is stronger than the interference (Hu and Wang, 2004). The system of Hu and Wang (2004) is a voiced-speech segregation system that utilizes differences in periodicity between target and interference. The similarities between target and interference characteristics degrade the performance of the CASA system and therefore contribute to informational masking in our model.

Lippmann and Carlson (1997) used *a priori* reliabilities of T-F regions to examine the effects of noise, low- and high-pass filtering, and interruptions by periodic silent gaps on the performance of a missing-data recognizer on a digit recognition task. This is then compared to human performance on a consonant-vowel-consonant syllable recognition task. Cooke (2006) also used a missing-data recognizer to model the effects of energetic masking on listeners' perception in babble noise. His model uses *a priori* knowledge of target-dominant T-F regions or "glimpses." In addition to

[a)]Present address: Robert Bosch LLC, Research and Technology Center North America, Pittsburgh, PA 15212. Electronic mails: srinivasan.36@osu.edu and soundar.srinivasan@us.bosch.com
[b)]Electronic mail: dwang@cse.ohio-state.edu

modeling energetic masking, Barker and Cooke (2004) modeled informational masking by performing grouping using trained speech models in a top-down manner (see Barker *et al.*, 2005). The present study addresses the problem of modeling the effects of both energetic and informational masking on multitalker speech perception by combining speech segregation based on *a priori* pitch and missing-data recognition (see Srinivasan and Wang, 2005a for an earlier version).

The model proposed here could also serve as an architecture for robust speech recognition in the presence of interfering speech sources. It is well known that the performance of automatic speech recognizers (ASRs) degrades rapidly in the presence of other sound sources (Huang *et al.*, 2001; Srinivasan, 2006). Speech recognizers are typically trained in an environment containing a single speech source and face a problem of mismatch when used in conditions where target speech occurs simultaneously with other sources. To mitigate the effect of this mismatch on recognition, "noisy" speech is typically preprocessed by speech separation systems. However, in many realistic applications, the output of typical speech segregation algorithms contains distortions in segregated speech not seen during ASR training. These distortions cause substantial degradation in recognition performance (Cooke *et al.*, 2001). Researchers have previously shown that combining binaural speech segregation with missing-data methods can improve target speech recognition under multitalker conditions (Roman *et al.*, 2003; Palomaki *et al.*, 2004). Unlike these binaural systems, the model presented here combines monaural target segregation and missing-data recognition. Our evaluations show that the proposed model can improve robust speech recognition under monaural multitalker conditions.

Models such as the articulation index (Fletcher, 1953) and the speech-transmission index (Steeneken and Houtgast, 1980) predict speech intelligibility in the presence of noise and other certain distortions. As pointed out by Cooke (2006), the use of an ASR system as a component in a model of human speech recognition additionally enables it to predict listener responses on a per-utterance basis. The purpose of the present study is twofold. First, similar to Cooke (2006), the proposed system is used to model listener performance on a per-utterance basis using a common multitalker speech corpus. Second, the proposed model is intended to serve as a potential paradigm for automatic speech recognition in multitalker situations.

The rest of the paper is organized as follows. In the next section, we briefly review a recent study that systematically examined the degradation in speech perception caused by energetic and informational masking using a binary masking procedure (Chang, 2004; Brungart *et al.*, 2006). Section III contains a detailed presentation of our proposed model. The model has been systematically evaluated on the same task used in the perceptual study presented in Sec. II. The evaluation results and a comparison with listener performance are presented in Sec. IV. Finally, conclusions are given in Sec. V.

II. ENERGETIC AND INFORMATIONAL MASKING EFFECTS IN MULTITALKER SPEECH PERCEPTION

A recent study (Chang, 2004; Brungart *et al.*, 2006) uses binary T-F masks to isolate the effects of energetic and informational masking on the intelligibility of a target speech signal in the presence of one or more competing speech signals. Specifically, Brungart *et al.* (2006) utilized an ideal binary mask that is obtained from premixed target and interference as follows. A unit in the ideal binary mask is assigned a value 1 if the signal-to-noise ratio (SNR) within the corresponding T-F unit exceeds a predefined local SNR criterion (LC) value; it is labeled 0 otherwise. To generate this mask, target and interfering signals are first analyzed using a 128-channel gammatone filterbank whose center frequencies are quasilogarithmically spaced from 80 Hz to 5 kHz (Patterson *et al.*, 1988). The energy at the output of each filter is calculated every frame for both target and interference. Each rectangular frame is 20 ms long with a 10 ms frame shift (a frame rate of 100 Hz). The ideal binary mask is used to resynthesize a signal by retaining only those T-F units in the mixture where the local SNR exceeds the specified LC value. Specifically, the gammatone filter responses of a mixture are weighted by the binary mask and summed across frequencies (after accounting for across-filter phase shifts) to yield a resynthesized signal. The resynthesized signal is then used in a series of experiments to study the effects of the number of interfering talkers and their sex on the two types of masking.

The speech corpus used by Chang (2004) and Brungart *et al.* (2006) is the coordinate response measure (CRM) corpus (Bolia *et al.*, 2000). This corpus consists of utterances from four male and four female speakers produced according to the grammar, "READY ⟨$call-sign⟩ GO TO ⟨$color⟩ ⟨$digit⟩ NOW." There are eight call signs, four colors, and eight numbers (Bolia *et al.*, 2000) and the target utterance always contains the call sign. "BARON;" e.g., "READY BARON GO TO RED ONE NOW." The interference utterance consists of a call sign, a color, and a number different from that of the target. Figure 1 shows the effect of applying the ideal binary mask to a mixture of two speech utterances from this corpus. Figures 1(a) and 1(b) show the cochleagrams of a target speech utterance and an interference utterance, respectively. A cochleagram is a T-F representation of a signal analogous to a spectrogram and is generated using the gammatone filterbank decomposition of a signal as described before (see Wang and Brown, 2006). The target signal corresponds to a male speech utterance, "READY BARON GO TO BLUE ONE NOW." The interference corresponds to a female utterance, "READY ARROW GO TO RED THREE NOW." Figure 1(d) shows the cochleagram of a mixture of target and interference at 0 dB SNR. The ideal binary mask for this mixture corresponding to a LC value of 0 dB is shown in Fig. 1(c). T-F units labeled 1 in the mask are shown in black and white represents the T-F units labeled 0. This mask is applied to the mixture in Fig. 1(d) and the results are presented in Fig. 1(e). Note that the application of the binary mask results in the removal of the interference-dominant T-F units from the mixture.

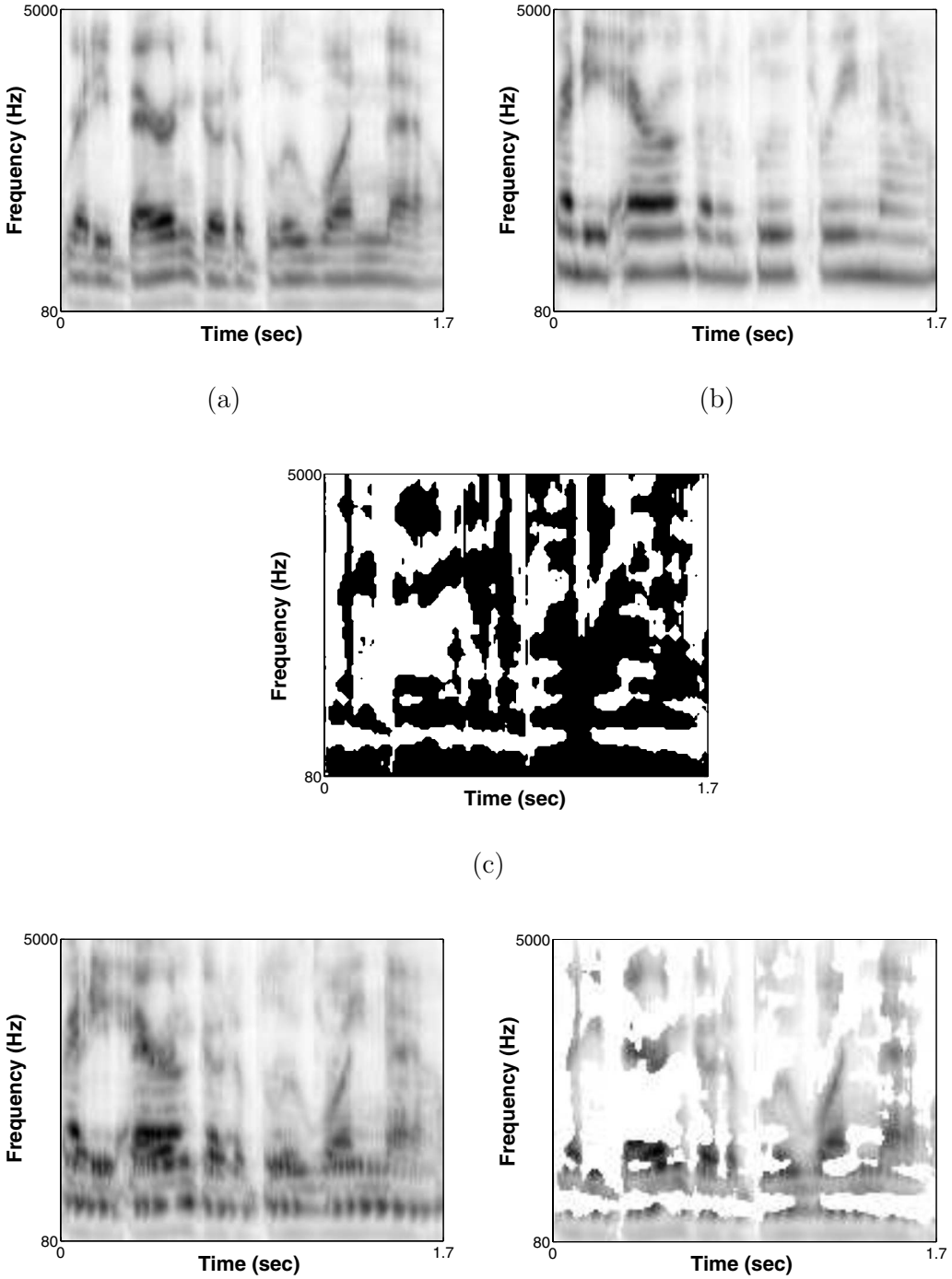


FIG. 1. An illustration of the ideal binary mask for a mixture of a male target utterance and a female interference. (a) The cochleagram of the male target utterance. (b) The cochleagram of the female interference utterance. (c) An ideal binary mask at 0 dB LC value. The target-dominant T-F units are marked black and the interference-dominant T-F units are marked white. (d) The cochleagram of the mixture. (e) The cochleagram obtained from (d) by applying the ideal mask in (c).

Using signals resynthesized by applying such ideal binary masks to multitalker speech mixtures, Brungart *et al.* (2006) asked listeners to identify the keywords (the color and the number) in the target phrase. The target-to-masker ratio (TMR) in the mixtures was fixed at 0 dB. While SNR refers to the ratio of target energy and combined interference energy, TMR refers to the ratio of target energy and energy of one interference in the mixture (interference signals have equal energy) (Brungart *et al.*, 2001). Figure 2 shows the effects of varying the number of competing talkers on the correct identification of keywords in the target phrase as a function of the LC value. The results are shown in terms of accuracy of recognizing both the color and the number in the target phrase at different LC values, ranging from −60 to +30 dB in steps of 3 dB. A control, "no mask" condition, is also included to assess listener performance directly on the

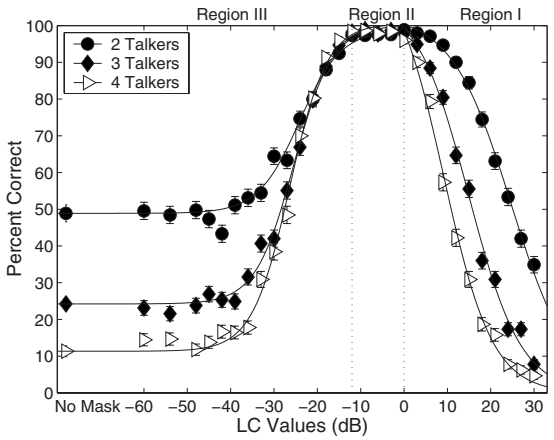


FIG. 2. Percentage of trials in which the listeners identified the keywords in the target phrase correctly (from Brungart *et al.*, 2006). The error bars represent 95% confidence intervals. The figure shows the effect of varying the number of competing talkers on listener performance as the LC value is varied in steps of 3 dB.

mixture in each multitalker condition. In other words, the signals generated in this control condition did not undergo any binary mask processing.

Based on the results in Fig. 2, Brungart *et al.* (2006) reached the following main conclusion. The information lost in the target signal due to energetic masking is argued to be approximately the same as that resulting from the resynthesis using an ideal binary mask with a LC value of 0 dB. Note that this mask removes all interference-dominant T-F units and, despite distortions caused by the removal of certain T-F regions of the mixture, the resulting intelligibility is quite high. Brungart *et al.* (2006) also suggested that the signal resynthesized from an ideal binary mask with a positive LC value of δ dB (region I) causes energetic masking equivalent to that by a mixture with SNR of $-\delta$ dB [see also Fig. 9(a)]. Therefore, the decrease in listener performance in region I as LC increases shows the increase in the deleterious effects of energetic masking. On the other hand, decreasing the LC value below 0 dB allows for the progressive introduction of interference energy in the resynthesized signal. This leads to an increase in the level of informational masking. Hence, energetic masking effects dominate listener perception in region I, while informational masking effects dominate in region III of Fig. 2. The signals employed in region II include some T-F units dominated by interference. However, listener performance seems to be largely unaffected by the interference contained in those units. Incidentally, region II, where near perfect performance is obtained, is centered at a LC value of −6 dB.

III. A COMPUTATIONAL MODEL

Listener performance gradually degrades as the LC value decreases in region III of Fig. 2. We propose to attribute this degradation to performance limitations of ASA. In other words, listeners use differences in the voice characteristics of target and interfering talkers to segregate the target speaker with varying degrees of success. Additionally, region II demonstrates the robust performance of listeners in the presence of energetic masking as isolated by the use of ideal binary masks. These observations motivate our model for multitalker speech perception shown in Fig. 3. The proposed model is a two-stage system that combines a CASA system with a missing-data ASR. The input to the model in Fig. 3 is a mixture of target and interference, sampled at 20 kHz. We use the same auditory filterbank decomposition of the input signal as used by Brungart *et al.* (2006) (see Sec. II). The output is used to generate feature vectors for recognition and as input to a monaural CASA system (Sec. III A). In our model, the monaural CASA system of Hu and Wang (2004) is adapted to segregate target from interference. The computational goal of the CASA system is an ideal binary mask (Wang, 2005). Human perception in the presence of energetic masking is modeled using a missing-data ASR that treats the masked T-F regions as missing. The similarities between target and interference cause, however, deviations in the estimation of the ideal binary mask by the CASA system. This corresponds to simulated informational masking in our model.

Figure 4 illustrates the effect of errors in the estimation of the ideal binary mask. Figure 4(a) shows the cochleagram of the same mixture shown in Fig. 1(d). Figure 4(b) shows the ideal binary T-F mask generated at the LC value of 0 dB. As mentioned before, this mask removes all interference-dominant T-F units. Figure 4(c) shows an estimated binary mask produced by the CASA system. Due to errors in target segregation, this mask contains many interference-dominant T-F units. Application of this mask to the cochleagram in Fig. 4(a), therefore, retains some interference energy in the masked mixture as shown in Fig. 4(e). Figure 4(d) shows the cochleagram obtained from (a) by applying the ideal mask in

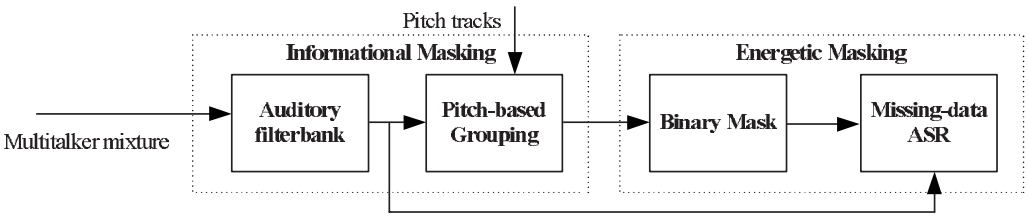


FIG. 3. Schematic of the proposed model. The input mixture signal is analyzed by an auditory filterbank in successive time frames. The output is fed to a monaural CASA system that uses pitch tracks of the individual sources in the mixture to produce a binary mask that selects the T-F regions in the mixture where target dominates interference. This mask is used by the missing-data recognizer to decode the input.

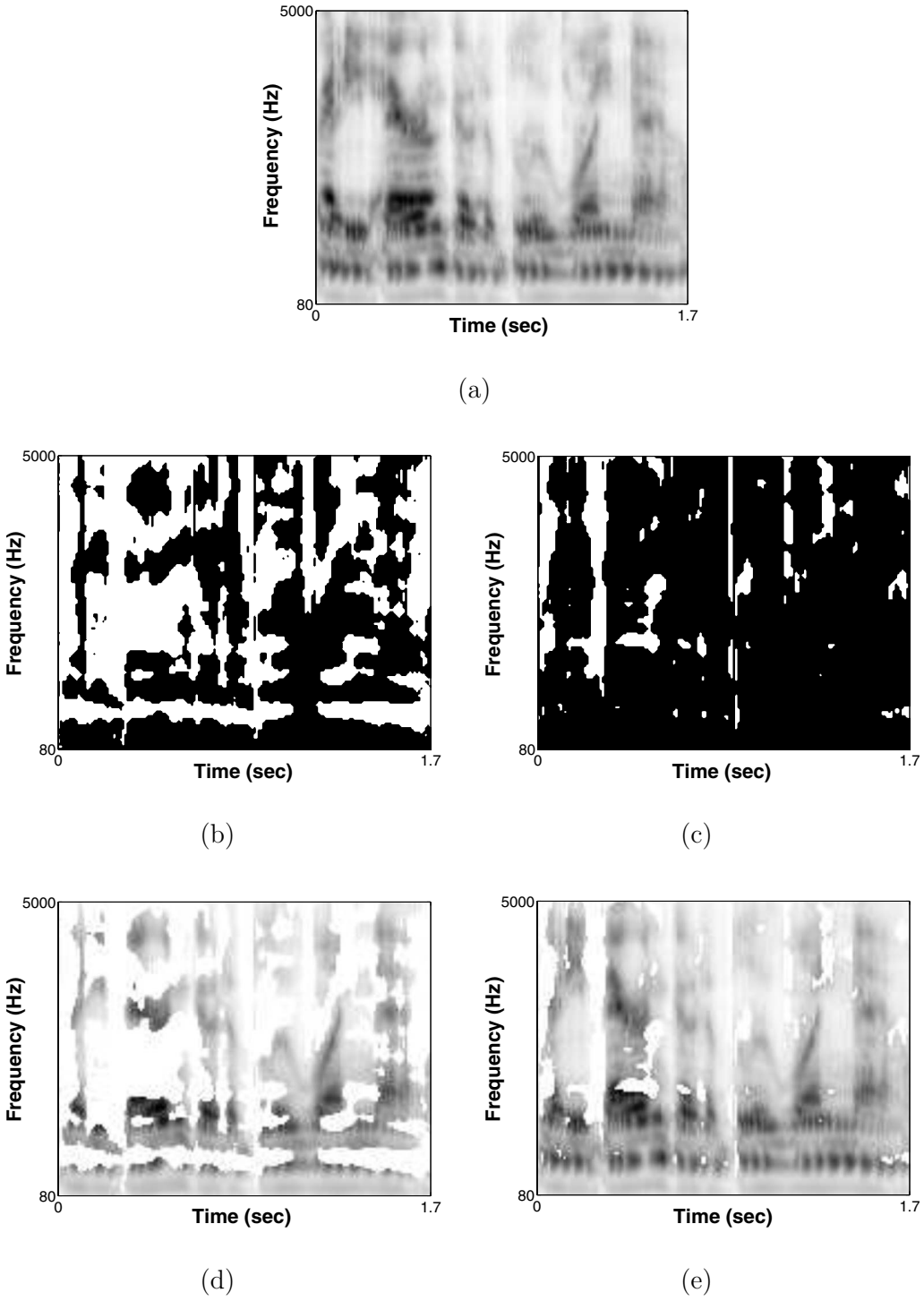


FIG. 4. An illustration of informational masking caused by deviations in target segregation via binary mask estimation. (a) The cochleagram of the mixture. (b) The ideal binary mask at 0 dB LC value. (c) An estimated binary mask. The target-dominant T-F units in (b) and (c) are marked black and the interference-dominant T-F units are marked white. (d) The cochleagram obtained from (a) by applying the ideal mask in (b). (e) The cochleagram obtained from (a) by applying the estimated mask in (c).

(b). The degradation in the model performance when using the cochleagram in Fig. 4(e) compared to the one in Fig. 4(d) is attributed to informational masking.

A. Pitch-based segregation of target

Under monaural conditions, the human auditory system is believed to segregate a target speech source from various interferences using several primitive cues, including differences in pitch (Bird and Darwin, 1997; Bregman, 1990; Brokx and Nooteboom, 1982) and onset, as well as prior knowledge (Bregman, 1990). Pitch has been successfully used for segregation of voiced speech in many CASA systems (Brown and Cooke, 1994; Hu and Wang, 2004). Here, we adapt the speech separation system by Hu and Wang (2004) to segregate target from interference. This system is chosen as it shows robust performance when tested with a

variety of intrusions. This system has two main stages: segmentation and grouping. In segmentation, the input signal is decomposed into a collection of contiguous T-F regions that are dominated by only one sound source. During grouping, the segments that likely belong to the same source are grouped together.

Consistent with psychophysical studies (Bird and Darwin, 1997), the system of Hu and Wang (2004) applies different processing strategies in the low- and high-frequency ranges. It has been shown that when neighboring channels respond to the same source, their autocorrelation responses are similar (Brown and Cooke, 1994; Wang and Brown, 1999). Therefore, cross correlation between adjacent channels indicates whether the filter channels respond to the same source. Hence, in the low-frequency range ($<$800 Hz), the system generates segments based on temporal continuity and cross-channel correlation. At higher frequencies, due to its wider bandwidth relative to harmonic spacing, a gammatone filter responds to multiple harmonics. Hence, the system uses the envelope characteristics of gammatone filter responses in the high-frequency range. Specifically, the cross-channel correlation of envelopes is used for segmentation along with temporal continuity.

For grouping segments, the system uses similarity in periodicity. The autocorrelation of a filter response in a frame encodes the periodicity within the corresponding T-F unit. Hence, if in a T-F unit the maximum autocorrelation value within the plausible pitch range has a lag consistent with the pitch lag of the target source in that frame, it is labeled as target-dominant (or 1); it is labeled interference dominant (or 0) otherwise (Hu and Wang, 2004). Furthermore, at high frequencies, a response envelope fluctuates at a rate consistent with the frequency of the dominant pitch, and amplitude modulation rates are therefore used for grouping (see also Cooke, 1993). To illustrate the potential of the proposed approach for segregation, pitch tracks and pitch strengths are derived *a priori* from premixed target and interfering signals using PRAAT (Boersma and Weenink, 2002). Note that robust multipitch tracking of two or more sources is currently a challenging problem (Wang and Brown, 2006).

Hu and Wang (2004) utilized only target pitch contours for grouping. However, listeners appear to utilize the pitch information of interfering sources too (de Cheveigne, 1997). In particular, the results of Culling *et al.* (2005) suggest that the auditory system can "perceptually remove" the harmonic components of at least one interference. Therefore, we adapt the system of Hu and Wang (2004) by expanding the grouping procedure as follows. First, we use the system of Hu and Wang (2004) to group segments based on the dominant pitch in a time frame. The dominant pitch in a frame is the one that has the highest associated pitch strength, which is measured as the height of the normalized autocorrelation value at the pitch lag. Second, the identity of the dominant pitch is then used to derive the final binary T-F mask in the following fashion. If the dominant pitch at a particular frame belongs to the interferer, we discard the grouped segments in that frame. Specifically, the grouped segments in that frame are set to 0 in the mask and the rest of the T-F units are set to 1 (see also de Cheveigne, 2005 for alternate approaches). On the other hand, if the dominant pitch belongs to the target, the grouped T-F units are retained. In this case, the grouped T-F units are labeled 1 in the mask, and other T-F units in that frame are labeled 0. We find that the proposed system performs better than the original system of Hu and Wang (2004). The performance improvement is especially significant when the original mixture SNR is high. In this case, canceling harmonic portions of interference helps retain more target-dominant T-F units than grouping harmonic portions of target alone. Figure 5 shows a comparison between the proposed system and the system of Hu and Wang (2004). The results are generated using the same target and interference in Figs. 1(a) and 1(b), mixed at a SNR of 20 dB. Figure 5(a) shows the binary mask containing the segregated voiced portions by primarily canceling the interference-dominant T-F units. The segregated T-F units are shown in black. All other T-F units are shown in white. Figure 5(b) shows the mask generated by the system of Hu and Wang (2004). For comparison, the ideal binary mask corresponding to this mixture, generated at 0 dB LC value, is shown in Fig. 5(c). Notice that the system of Hu and Wang (2004) wrongly labels some target-dominant T-F units as 0.

Figure 6 shows how the binary mask corresponding to the target is generated by using pitch strengths of both target and interference. Figure 6(a) shows a binary mask containing the segregated voiced portions corresponding to the dominant pitch in each frame. The segregated T-F units are labeled 1 and shown in black. All other T-F units are labeled 0 and shown in white. This mask is generated by using the mixture in Fig. 1(d) as input to the proposed CASA system. The dominant pitch in a frame may correspond to either target or interference. Figure 6(b) shows how pitch strengths of target and interference vary in the mixture. If the pitch strength of the interference is higher in a frame, mask labels are flipped; they are retained otherwise. The resulting mask is shown in Fig. 6(c). Currently, we do not process unvoiced frames. For such frames, all frequency units are labeled 0. The output of the CASA system in the form of a binary mask is then used by the missing-data ASR to recognize target speech.

B. Missing-data recognition

The feature vectors for the missing-data recognizer consist of the instantaneous Hilbert envelope at the output of each gammatone filter, smoothed using a first-order filter with 8 ms time constant and log compressed as suggested by Cooke *et al.* (2001). The missing-data recognizer is an HMM-based ASR that makes use of spectrotemporal redundancy in speech to recognize a noisy signal based on its target-dominant T-F units. Given a speech observation sequence X, the problem of word recognition is to maximize the posterior $P(W|X)$, where W is a valid word sequence. When parts of X are masked by noise or other distortions, a binary T-F mask can be used to partition X into its reliable and unreliable constituents as X_r and X_u, where $X=X_r \cup X_u$. The missing-data ASR treats the T-F regions labeled 0 as unreliable data during recognition. Specifically, it modifies the computation of the observation probability in a state of

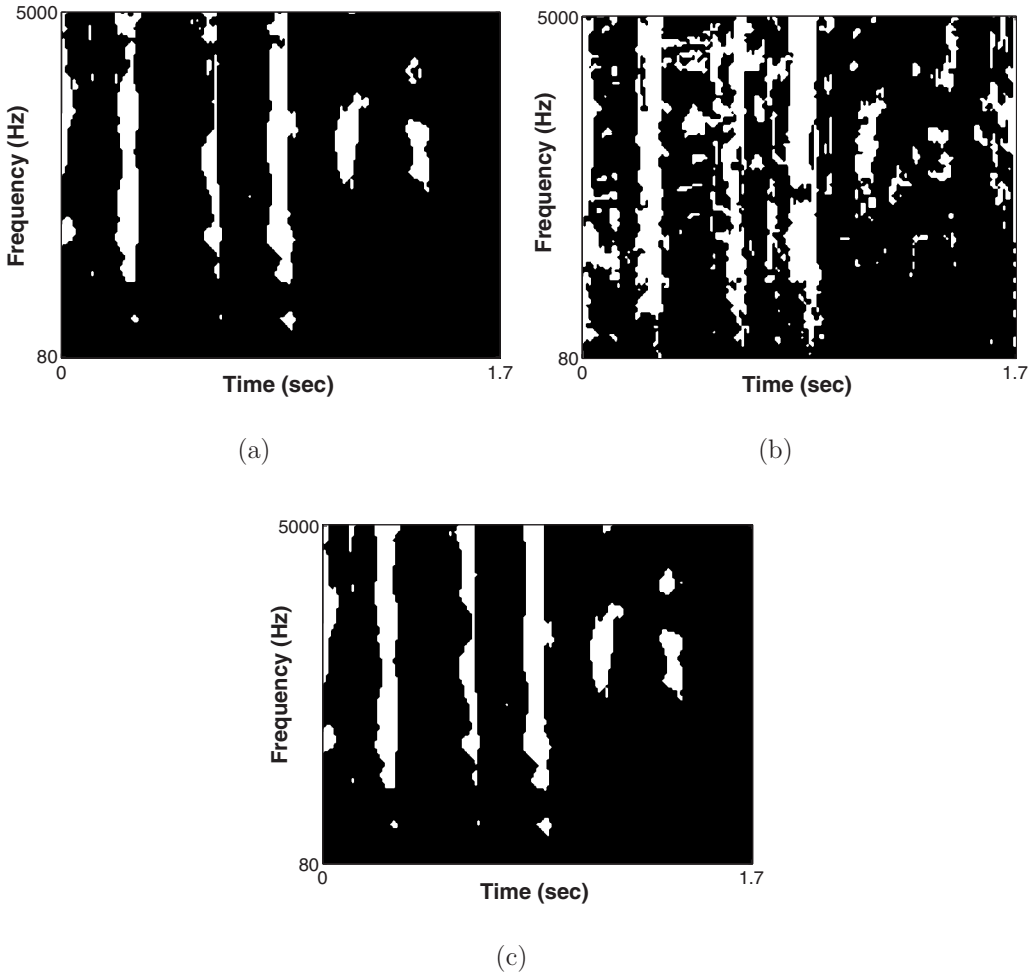


FIG. 5. An illustration of improved segregation using the proposed approach. (a). The binary mask generated by the proposed system. T-F units determined to be target-dominant are shown in black. T-F units consistent with the interference pitch are canceled and shown in white. (b) The binary mask produced by the system of Hu and Wang (2004). (c) The ideal binary mask.

an HMM (hidden Markov model) based ASR to handle missing or unreliable data (Cooke *et al.*, 2001). The observation density in an ASR is typically modeled using a mixture of Gaussians as follows:

$$p(x|q) = \sum_{k=1}^{M} p(k|q)p(x|k,q), \tag{1}$$

where x is the spectral energy feature vector in a frame, M is the number of mixture components, k is the mixture index, q is an HMM state, $p(k|q)$ is the mixture weight, and $p(x|k,q)=\mathcal{N}(x;\mu_{k,q},\Sigma_{k,q})$. Note that $\mathcal{N}(x;\mu_{k,q},\Sigma_{k,q})$ denotes that x follows a normal (Gaussian) distribution with mean $\mu_{k,q}$ and variance $\Sigma_{k,q}$. When parts of x are corrupted by interference, the missing-data ASR marginalizes unreliable feature dimensions in the computation of the likelihood in Eq. (1). Typically, the various dimensions of the feature vectors are modeled as independent given a mixture. Theoretically, this is a good approximation if an adequate number of mixtures are used (McLachlan and Basford, 1988). Hence, in the presence of unreliable data, the computation of the observation density is modified as

$$p(x|q) = \sum_{k=1}^{M} p(k|q) \prod_{j} p(x_{r,j}|k,q) \prod_{i} \int p(x_{u,i}|k,q)dx_{u,i}, \tag{2}$$

where $x_{r,j}$ and $x_{u,i}$ correspond to the spectral energies in a reliable (j) and an unreliable (i) feature dimension, respectively. Note that $p(x_{r,j}|k,q)=\mathcal{N}(x_{r,j};\mu_{k,q,j},\sigma^2_{k,q,j})$.

Furthermore, if the range for the true value of the unreliable feature is known, it provides bounds (limits) over which the unreliable feature is integrated. Under additive interference conditions, the true speech value $\tilde{x}_{u,i}$, in the unreliable part, may be constrained as $0 \leq \tilde{x}_{u,i} \leq y_{u,i}$ (Cooke *et al.*, 2001; Srinivasan and Wang, 2007), where $y_{u,i}$ is the observed (mixture) spectral energy. This constraint is then used as bounds on the integral in Eq. (2) as

$$p(x|q) = \sum_{k=1}^{M} p(k|q) \prod_{j} p(x_{r,j}|k,q) \prod_{i} \int_{0}^{y_{u,i}} p(x_{u,i}|k,q)dx_{u,i}. \tag{3}$$

This bounded marginalization method is shown in Cooke *et al.* (2001) to have a better recognition score than the

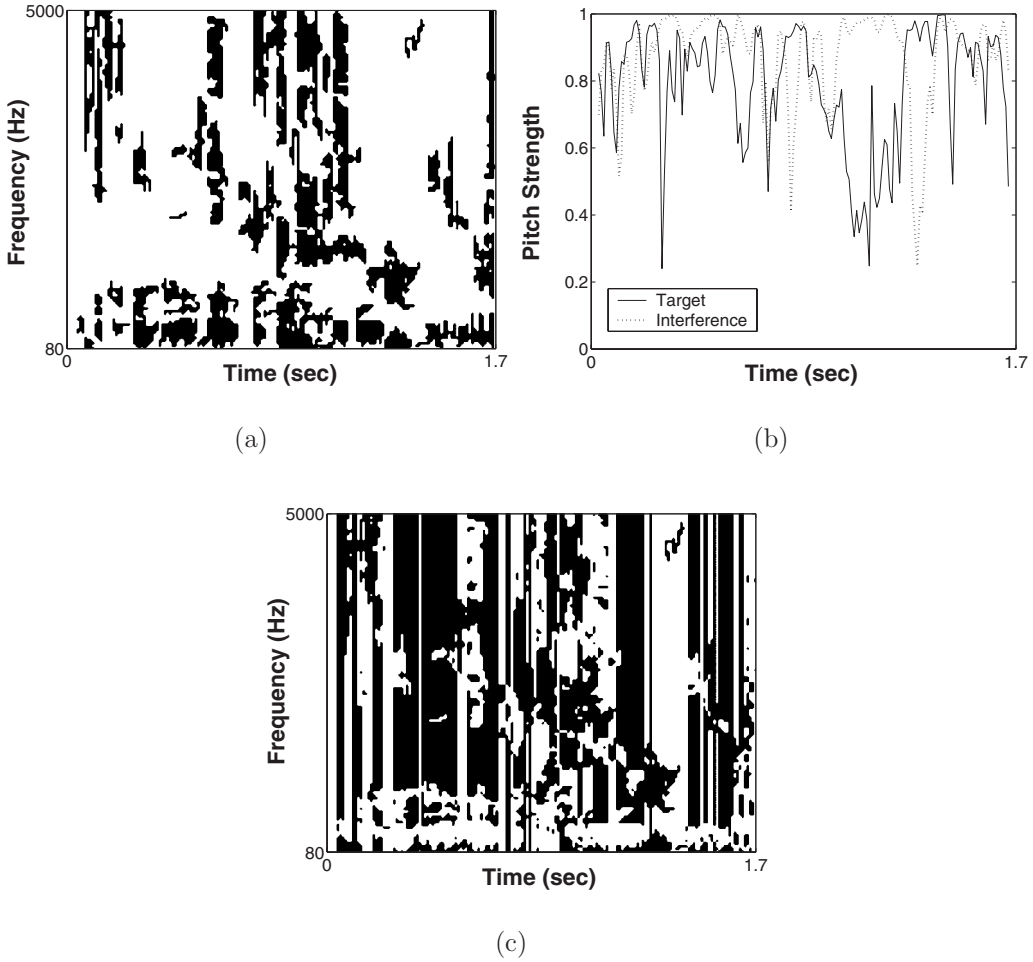


FIG. 6. An illustration of target segregation based on the dominant pitch. (a) The binary mask generated by the CASA system using the pitch cue. The input to the system is the mixture shown in Fig. 1(d). If the periodicity of a gammatone filter response is consistent with the dominant pitch in that frame, it is shown in black. All other T-F units are shown in white. (b) The variation in the strengths of target and interference pitch tracks across time. (c) The final mask is obtained from (a) by flipping the frame labels generated using the interference pitch.

simple marginalization method, and is hence used in all our experiments.

IV. EVALUATION RESULTS

To facilitate comparison with the behavioral data from Chang (2004) and Brungart *et al.* (2006), we have also evaluated our model using the CRM corpus (Bolia *et al.*, 2000). Twenty-three ("Ready," "Baron," "Charlie," "Arrow," "Laker," "Hopper," "Ringo," "Tiger," "Eagle," "Goto," "Blue," "Green," "Red," "White," "Now," and the numbers 1–8) speaker-independent, word-level HMMs are trained. All models have eight states, whose output distribution is modeled as a mixture of ten Gaussians with diagonal covariance. The models are trained using 1792 utterances from 3 male talkers (talkers 1–3) and 4 female talkers (talkers 4–7) in this database chosen arbitrarily. The testing data consist of utterances from a male talker (talker 0) not utilized during training and containing the call sign "BARON." Similar to the experiments by Brungart *et al.* (2006), one, two, and three utterances from the same talker are added to target as interference at 0 dB TMR. Thus the SNRs corresponding to two, three, and four talker mixture conditions are 0, −3, and −4.8 dB, respectively. Interference utterances contain call signs, numbers, and colors different from the target one. Each testing condition comprises 256 mixture utterances. An HMM toolkit, HTK (Young *et al.*, 2000), is used for training. For testing, a decoder incorporating the missing-data method is used. The task is to recognize both color and number in the

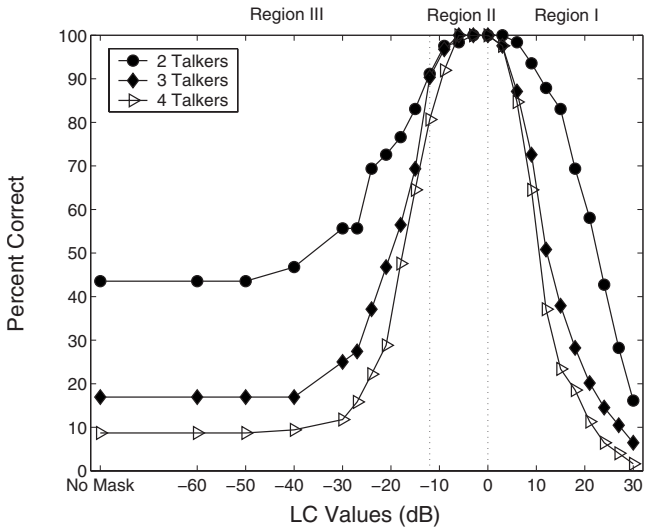


FIG. 7. Percentage of utterances in which the model identified the keywords in the target utterance correctly. The figure shows the model's performance with respect to the LC value under various multitalker conditions.

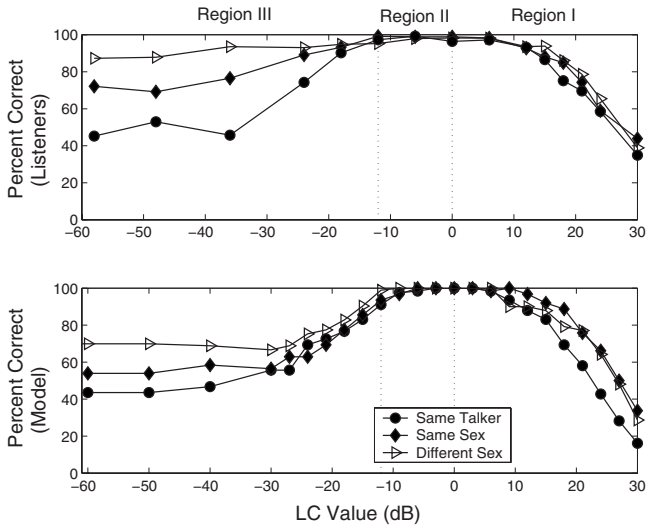


FIG. 8. The effect of voice characteristics on listener and model performance with two talkers as a function of the LC value. The top panel shows the performance of listeners in correctly identifying the keywords in the target utterance (from Chang, 2004). The bottom panel shows the performance of the proposed model on the same task.

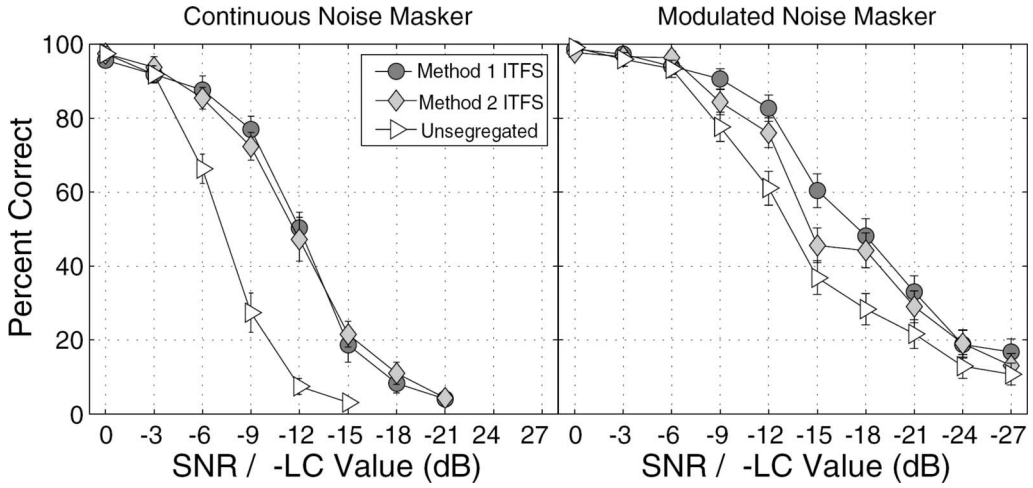
target utterance as described in Sec. II. We recognize the contents of the target utterance in the mixture by using the following grammar: "READY BARON GOTO ⟨$color⟩ ⟨$digit⟩ NOW." Recognition performance on the target-only test-data input is 100%. Note that the chance performance on this task is 3.125%.

Similar to the procedure employed by Brungart *et al.* (2006), the multitalker mixture is resynthesized using ideal binary masks with varying LC values (see also Sec. II). The resynthesized signal is used as input to the proposed model. Figure 7 shows the performance of the proposed model as a function of the LC value under different multitalker conditions. Similar to the evaluation used by Brungart *et al.* (2006), performance in Fig. 7 is shown in terms of percentage of utterances in which the model correctly identified all the keywords in the target phrase.

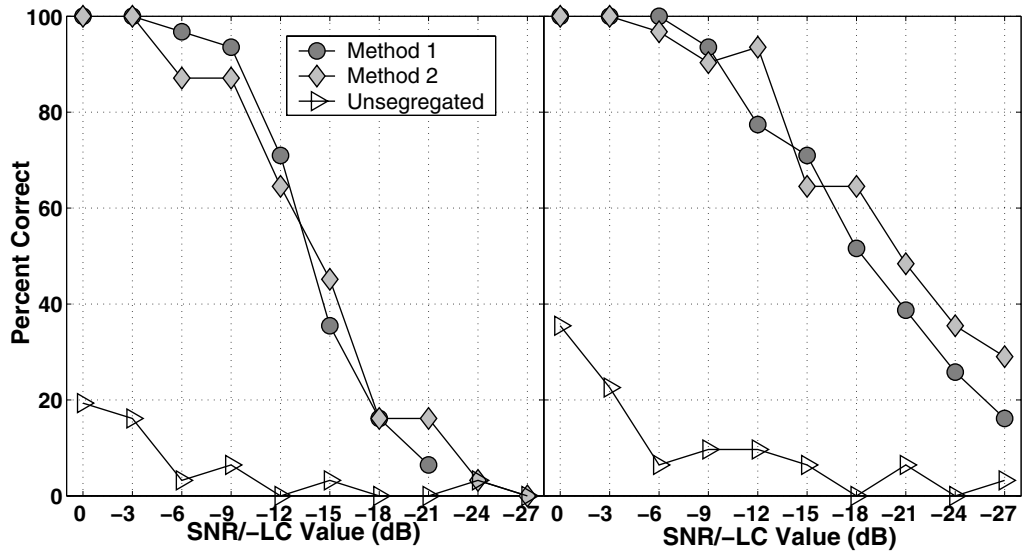
While the absolute recognition rates at some LC values differ from human performance, the model is able to simulate the general pattern of systematic listener performance seen in Fig. 2. As in Fig. 2, informational masking can be seen to have a larger effect on the recognition performance. The model saturates to the ceiling performance in region II. The improvement compared to the no mask condition shows that the removal of interference-dominant T-F units from the input contributes to substantial performance gains. As in Fig. 2, the peak performance is obtained in an interval of LC values approximately around −6 dB. As the number of talkers in the mixture increases, both model and human performance significantly degrade in region III. In comparison, smaller differences are seen in region I across the different multitalker conditions. Overall, the comparison with Fig. 2 shows that the proposed model is able to replicate listener perception under various multitalker conditions in the presence of both energetic (region I and parts of region II) and informational masking (region III).

Figure 8 shows the results of a second experiment in which the effects of voice characteristics of an interfering talker are examined. The top panel in Fig. 8 shows listener performance under three different interference conditions: Same talker, a different talker of the same sex, and a different sex talker (Chang, 2004). The bottom panel shows the performance of our model. In evaluating the model performance under the "same sex" condition, interference utterances are chosen randomly from talkers 1 to 3. For the "different sex" condition, interference utterances are chosen from talkers 4 to 7. The results of evaluation of our model and the human data show that the differences in voice characteristics between target and interference have only minor effects on the performance in regions I and II. The peak performance in both cases exhibits a plateau in region II. Recall that the SNR is the same across the three interference conditions. However, as with human listeners, the proposed model is able to utilize the larger differences in voice characteristics to obtain improved performance under the same sex and the different sex conditions in region III. Also, the model performance in the different sex condition is better that in the same sex condition. These indicate that the model is able to simulate the dependency of informational masking on the similarities in voice characteristics of target and interference.

Brungart *et al.* (2006) additionally examined the perception of a target speech utterance in the presence of nonspeech interferences. In this experiment, they considered two interfering sources: a continuous noise and a modulated noise. The continuous noise was generated by filtering a Gaussian noise source with the average long-term spectrum of all utterances in the CRM corpus. The modulated noise source was generated by further modulating the continuous noise masker with the envelope of a randomly chosen utterance in the CRM corpus (Brungart *et al.*, 2006). The resulting listener performance at various SNR and LC values is shown in Fig. 9(a). In "method 1," the SNR is fixed at 0 dB and the LC value is varied in steps of 3 dB. In "method 2," the LC value is kept constant at 0 dB and the SNR is varied from 0 to 27 dB in steps of 3 dB. Additionally, Brungart *et al.* (2006) performed a control experiment in which the listener performance was assessed without any binary mask processing at the aforementioned SNRs (equivalent to the no mask condition described in Sec. II), referred to as "unsegregated" in Fig. 9(a). The results in Fig. 9(a) were also used to validate the assumption that a δ dB increase in the LC value at a fixed SNR in region I in Fig. 2 is equivalent to a decrease in mixture SNR of $-\delta$ dB at a fixed LC value. Figure 9(b) shows the model performance in the continuous and the modulated noise conditions. Note that for both model and listeners, the performances in method 1 and method 2 conditions are quite similar. While the model performance in the unsegregated condition is well below the listener performance, the model is able to simulate the increased intelligibility in the modulated noise condition compared to that in the continuous noise condition. Additionally, the model performance is still better than a direct recognition of mixture speech. For example, at 0 dB SNR, the model achieves 16% and 17% absolute improvements in accuracy over the baseline recognition performance. The mask estimated by the proposed CASA system incorrectly labels many interference-



(a)



(b)

FIG. 9. The effect of noise characteristics on listener and model performance. For method 1, the performance is shown as a function of the LC value, while for method 2 and the unsegregated conditions, the performance is shown with respect to SNR. (a) The figure shows the performance of listeners in correctly identifying the keywords in the continuous and modulated noise conditions (from Brungart *et al.*, 2006). (b) The corresponding performance of the proposed model.

dominant T-F units as reliable, causing insufficient target segregation. This is the cause of the performance gap (compared to human listeners) in the unsegregated condition.

As described in Sec. I, the second objective of the present study is to examine the efficacy of the proposed model for robust speech recognition under multitalker conditions. In Fig. 7 and in the bottom panel of Fig. 8, the no mask condition represents a control condition in which the mixture signal is used directly as input to the model. The model performance with this input, therefore, shows the improvement over the baseline recognition performance. Table I shows the performance of the proposed model in the no mask condition. Performance is again reported in terms of the percentage of utterances in which the model correctly identified both keywords spoken by the target speaker. As mentioned before, the interference consists of one, two, and three utterances from the same talker as target at a TMR of 0 dB. Table I also shows the baseline performance obtained by using the mixture directly as input to the ASR. Note that, by performing segregation based on *a priori* pitch, the model is able to improve significantly over the baseline performance across all multitalker conditions. The performance

TABLE I. Keyword recognition accuracy (%) of the proposed model with respect to the number of talkers in the mixture. For comparison, the baseline accuracy is also shown.

No. of talkers	Baseline performance	Model performance
2	15.6	43.6
3	6.2	16.9
4	4.3	8.7

TABLE II. Keyword recognition accuracy (%) of the proposed model across various interference conditions. For comparion, the baseline accuracy is also shown.

Interference type	Baseline performance	Model performance
Same talker	15.6	43.6
Same sex	17.9	52.3
Different sex	21.2	69.9

improvement is especially substantial in the two talker condition. This is because the SNR in the two talker condition is higher. Also, with additional talkers, the mixture becomes more similar to babble and segregation based on dominant pitch deteriorates.

Similarly, Table II compares the model and the baseline performance under same talker, same sex and different sex conditions. There are two talkers in the mixture at an SNR of 0 dB. The pitch-based grouping component of the model is able to utilize the larger differences between target and interference pitch contours under same sex and different sex conditions to produce better segregation results. This leads to a substantial improvement in the performance of the missing-data recognizer. In contrast, the baseline performance only changes slightly under those conditions.

V. DISCUSSION

We have presented a model for monaural multitalker speech perception that is able to account for the effects of both energetic and informational masking in multitalker conditions. We have conducted a systematic comparison between the model performance and listener results on a common speech corpus and shown that the performance of the proposed model is in broad quantitative agreement with behavioral data. We have also shown that differences between target and interference pitch ranges contribute to a reduction in informational masking (see also Oh and Lutfi, 2000) by improving target speech segregation. Hence, the performance of the model is better in the same sex and the different sex conditions compared to that in the same talker condition. Notice that the best performance, for both human listeners and the proposed model, is obtained under the different sex condition. This is expected due to the difference in voice characteristics between target and interference being largest under this condition.

Our study also demonstrates that combining target speech segregation via CASA and missing-data recognition can provide significant improvement over baseline recognition performance. Hence, the proposed model shows potential for robust speech recognition under multitalker conditions. The present study, therefore, confirms previous findings that the harmonicity of voiced speech can be successfully exploited to estimate masks for missing-data recognition of monaural noisy speech mixtures (Seltzer *et al.*, 2000; Brown *et al.*, 2001; van Hamme, 2004). Additionally, our study extends previous studies to multitalker conditions.

The results of the present study could be used in conjunction with the findings of Brungart *et al.* (2006) and Chang (2004) to design appropriate binary masking thresholds for CASA systems. Note that CASA systems must estimate the ideal T-F mask directly from the mixture. As in the present study, the computational goal of most current CASA systems is an ideal binary mask generated an LC value of 0 dB. The results of Brungart *et al.* (2006) and Chang (2004) suggest that speech segregated by using a −6 dB LC value may be a better choice for improving speech intelligibility for human listeners. However, our results indicate that for missing-data recognition, this choice of LC value may only be appropriate in the two talker condition. For more than two talkers, the peak performance is centered at a value higher than −6 dB.

Our model utilizes only pitch information for grouping and obtains large improvements in recognition accuracy. The use of other ASA cues including common onset should help further enhance segregation, especially for unvoiced speech (Hu and Wang, 2005, 2007). These primitive grouping mechanisms may also be supplemented by schema-based, top-down segregation (Barker *et al.*, 2005; Srinivasan and Wang, 2005b, 2005c). For example, top-down processing may play an important role in segregation of unvoiced consonants. Top-down processing may also help in segregation of voiced speech when SNR is low. In general, top-down processing provides prior information for grouping, and in a complete system, top-down and bottom-up CASA should probably interact for maximal performance gains.

The model also needs to address sequential grouping of sources across time (Bregman, 1990). In our current work, we have avoided this problem by utilizing the *a priori* pitch information and the target grammar. While solutions based on speech or speaker models have been proposed (Barker *et al.*, 2005; Shao and Wang, 2006), such solutions are currently limited to mixtures of one talker and nonspeech interference or mixtures of two talkers. Future work needs to develop a general solution for arbitrary speech mixtures.

Note that for use as a robust speech recognition system, the main limitation is the assumption of *a priori* pitch tracks. While algorithms for forming multiple continuous pitch contours exist (Wu *et al.*, 2003), the problem of sequential grouping of pitch contours into pitch tracks has been little addressed and presents a major challenge for CASA systems. Recall that the performance improvements reported in this study come despite not segregating unvoiced speech. Segregation of unvoiced speech, as mentioned above, should further improve recognition results. Additional improvements can also be obtained by using binary T-F masks in an uncertainty decoding approach to robust speech recognition (Srinivasan and Wang, 2007).

ACKNOWLEDGMENTS

This research was supported in part by an AFOSR grant (FA9550-04-1-0117), an NSF grant (IIS-0534707), and an AFRL grant via Veridian. We thank P. S. Chang, D. S. Brungart, and L. L. Feth, for helpful discussions. A preliminary version of this work was presented in 2005 Interspeech.

Barker, J., and Cooke, M. P. (**2004**). "Modelling the intelligibility of multitalker speech in the CRM task," presented on the *International Conference on Auditory Scene Analysis and Speech Perception by Human and Ma-*

chine, Hanse Institute for Advanced Studies, Delmenhorst, Germany.
Barker, J. P., Cooke, M. P., and Ellis, D. P. W. (**2005**). "Decoding speech in the presence of other sources," Speech Commun. **45**, 5–25.
Bird, J., and Darwin, C. J. (**1997**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London, UK), pp. 263–269.
Boersma, P., and Weenink, D. (**2002**). "PRAAT: Doing phonetics by computer, version 4.0.26," http://www.fon.hum.uva.nl/praat (last viewed October, **2007**).
Bolia, R. S., Nelson, W. T., and Ericson, M. A. (**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).
Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.
Brown, G. J., Barker, J., and Wang, D. L. (**2001**). "A neural oscillator sound separator for missing data speech recognition," in Proceedings of the International Joint Conference on Neural Networks'01, pp. 2907–2912.
Brown, G. J., and Cooke, M. P. (**1994**). "Computational auditory scene analysis," Comput. Speech Lang. **8**, 297–336.
Brungart, D. S., 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.
Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.
Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.
Chang, P. S. (**2004**). "Exploration of behavioral, physiological, and computational approaches to auditory scene analysis," Master's thesis, Department of Computer Science & Engineering, The Ohio State University; http://www.cse.ohio-state.edu/pnl/theses.html (last viewed October, 2007).
Cooke, M. P. (**1993**). *Modeling Auditory Processing and Organization* (Cambridge University Press, Cambridge, UK).
Cooke, M. P. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.
Cooke, M. P., Green, P., Josifovski, L., and Vizinho, A. (**2001**). "Robust automatic speech recognition with missing and unreliable acoustic data," Speech Commun. **34**, 267–285.
Culling, J. F., Linsmith, G. M., and Caller, T. L. (**2005**). "Evidence for a cancellation mechanism in perceptual segregation by differences in fundamental frequency," J. Acoust. Soc. Am. **117**, 2600.
de Cheveigne, A. (**1997**). "Concurrent vowel identification. III. A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. **101**, 2857–2865.
de Cheveigne, A. (**2005**). "The cancellation principle in acoustic scene analysis," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Norwell, MA), pp. 245–259.
Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.
Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, Princeton, NJ).
Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.
Hu, G., and Wang, D. L. (**2004**). "Monaural speech segregation based on pitch tracking and amplitude modulation," IEEE Trans. Neural Netw. **15**, 1135–1150.
Hu, G., and Wang, D. L. (**2005**). "Separation of fricatives and affricates," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing'05, pp. 1101–1104.
Hu, G., and Wang, D. L. (**2007**). "Auditory segmentation based on onset and offset analysis," IEEE Trans. Audio, Speech, Lang. Process. **15**, 396–405.
Huang, X., Acero, A., and Hon, H. (**2001**). *Spoken Language Processing* (Prentice-Hall, Upper Saddle River, NJ).
Lippmann, R. P., and Carlson, B. A. (**1997**). "Using missing feature theory to actively select features for robust speech recognition with interruptions, filtering, and noise," in Proceedings of the European Conference on Speech Communication and Technology'97, pp. 37–40.
Mayer, A. M. (**1876**). "Research in acoustics," Philos. Mag. **2**, 500–507.
McLachlan, G. J., and Basford, K. E. (**1988**). *Mixture Models: Inference and Applications to Clustering* (Dekker, New York).
Oh, E. L., and Lutfi, R. A. (**2000**). "Effect of masker harmonicity on informational masking," J. Acoust. Soc. Am. **108**, 706–709.
Palomaki, K. J., Brown, G. J., and Wang, D. L. (**2004**). "A binaural processor for missing data speech recognition in the presence of noise and small-room reverberation," Speech Commun. **43**, 361–378.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). "An efficient auditory filterbank based on the gammatone function," MRC Applied Psychology Unit (APU) Report No. 2341, Cambridge, UK.
Pollack, I. (**1975**). "Auditory informational masking," J. Acoust. Soc. Am. **57**, Supplement 1, p. 55.
Roman, N., Wang, D. L., and Brown, G. J. (**2003**). "Speech segregation based on sound localization," J. Acoust. Soc. Am. **114**, 2236–2252.
Seltzer, M. L., Raj, B., and Stern, R. M. (**2000**). "Classifier-based mask estimation for missing feature methods of robust speech recognition," in Proceedings of the International Conference on Spoken Language Processing '00, pp. 538–541.
Shao, Y., and Wang, D. L. (**2006**). "Model-based sequential organization in cochannel speech," IEEE Trans. Audio, Speech, Lang. Process. **14**, 289–298.
Srinivasan, S. (**2006**). "Integrating computational auditory scene analysis and automatic speech recognition," Ph.D. thesis, Biomedical Engineering Department, The Ohio State University, Columbus, OH.
Srinivasan, S., and Wang, D. L. (**2005a**). "Modeling the perception of multitalker speech," in Proceedings of the Interspeech '05, pp. 1265–1268.
Srinivasan, S., and Wang, D. L. (**2005b**). "Robust speech recognition by integrating speech separation and hypothesis testing," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing '05, Vol. **1**, pp. 89–92.
Srinivasan, S., and Wang, D. L. (**2005c**). "A schema-based model for phonemic restoration," Speech Commun. **45**, 63–87.
Srinivasan, S., and Wang, D. L. (**2007**). "Transforming binary uncertainties for robust speech recognition," IEEE Trans. Audio, Speech, Lang. Process. **15**, 2130–2140.
Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.
Tanner, W. P., Jr. (**1958**). "What is masking?" J. Acoust. Soc. Am. **30**, 919–921.
van Hamme, H. (**2004**). "Robust speech recognition using cepstral domain missing data techniques and noisy masks," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing '04, Vol. **1**, pp. 213–216.
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. (**1999**). "Separation of speech from interfering sounds based on oscillatory correlation," IEEE Trans. Neural Netw. **10**, 684–697.
Wang, D. L., and Brown, G. J. (**2006**). *Computational Auditory Scene Analysis: Principles, Algorithms and Applications* (Wiley, New York, IEEE, Hoboken, NJ).
Watson, C. S. (**2005**). "Some comments on informational masking," Acta. Acust. Acust. **91**, 502–512.
Wu, M., Wang, D. L., and Brown, G. J. (**2003**). "A multipitch tracking algorithm for noisy speech," IEEE Trans. Speech Audio Process. **11**, 229–241.
Young, S., Kershaw, D., Odell, J., Valtchev, V., and Woodland, P. (**2000**). *The HTK Book (for HTK Version 3.0)* (Microsoft Corporation, Redmond, WA).

Study of an automated procedure for a Dutch sentence test for the measurement of the speech reception threshold in noise

Hayo Terband[a] and Rob Drullman
TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands

(Received 17 May 2007; revised 20 August 2008; accepted 27 August 2008)

A procedure was developed for the automated measurement of the speech reception threshold in stationary noise (SRTn), which can be administered by the subjects themselves using a computer. The procedure was based on the SRTn test for Dutch developed by Plomp and Mimpen [(1979). "Improving the reliability of testing the speech reception threshold for sentences," Audiology, **18**, 43–52]. Because in the automated procedure the responses were entered on a keyboard, the question of how to deal with typing and spelling errors played a key role. At first the possibility of scoring on keywords only was examined. An experiment was conducted in which the adaptive procedure was varied. Results showed that the combination of scoring each keyword separately and a fixed scheme of the adaptation of the signal-to-noise ratio throughout the procedure yields the highest test-retest reliability. Subsequently, the collection and verification of responses using a keyboard were examined. Two different algorithms were developed and evaluated against the traditional task of verbal repetition and response verification by an experimenter. The results indicated a preference for verification by dynamic alignment over a spelling checker approach. In conclusion, the results show that it is possible to automate the test procedure while maintaining sufficient reliability.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2990706]

PACS number(s): 43.71.Gv, 43.71.Ky, 43.71.Es [DOS] Pages: 3225–3234

I. INTRODUCTION

Listeners with hearing loss generally consider the understanding of speech in noisy environments to be the most troublesome aspect of their impairment (Wagener *et al.*, 2003; Smits *et al.*, 2004). The ability to understand speech in noise can be quantified as the signal-to-noise ratio (SNR) necessary to reach a 50% intelligibility score, which is denoted as the speech reception threshold in noise (SRTn) [e.g., see Plomp and Mimpen (1979), Letowski *et al.* (1992), and Noordhoek *et al.* (1999, 2000)]. To measure this ability, SRTn tests have been developed for a wide variety of languages. Most of these tests consist of a repetition task of a simple straightforward sentence material that is presented against a background of (stationary speech-shaped) noise. Although this testing method has proved to be generally very accurate and reliable [e.g., see Smoorenburg (1992), Hagerman and Kinnefors (1995), Kollmeier and Wesselkamp (1997), and Versfeld *et al.* (2000)] it has the disadvantage that an experimenter is needed to administer the test and to decide whether a sentence is repeated correctly or not. The elimination of this disadvantage requires an automatic test procedure, and several automatic tests have been developed. For example, the telephone screening test for Dutch, with simple three-digit stimuli, proved quite successful in its domain (Smits *et al.*, 2004). In the present paper, we present the development of a fully automated *sentence* test that subjects can take independently on a computer.

For Dutch, the test developed by Plomp and Mimpen (1979) is considered to be the standard (Smoorenburg, 1992; Versfeld *et al.*, 2000; Smits *et al.*, 2004), and this test is widely used for both clinical and research purposes [e.g., see Plomp (1986), Rooij and Plomp (1991), and Noordhoek *et al.* (1999, 2000)]. The Plomp and Mimpen test consists of a list of 13 sentences of eight to nine syllables that are presented in a background of speech-shaped noise. The noise level is fixed at a comfortable level. The level of the sentences is varied according to an adaptive procedure. A sentence is considered to be repeated correctly only if the whole sentence is reproduced verbatim. The first sentence of a list is presented at a SNR of −8 dB, which is normally below the reception threshold. This sentence is repeated, each time at a 4 dB higher SNR until the subject can reproduce it correctly. The remaining 12 sentences are each presented once in a simple up-down procedure with a 2 dB step size. The speech reception threshold is calculated by averaging the presentation levels of sentences 5–13 and the level at which a 14th sentence would have been presented (Plomp and Mimpen, 1979).

The goal of our study was to design an automatic procedure based on the existing speech material, with a test-retest reliability similar to the standard SRTn-test by Plomp and Mimpen (1979). The first question that arose when considering the automation of a SRTn-test was whether, and if so how, the test method needed to be adapted. It might not be possible to maintain the same test procedure, and automation might introduce factors that affect the test validity. On the other hand, automation might enable methodological adjustments that improve the test validity. In an automatic proce-

[a]Present address: ENT/Medical Psychology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

dure, the responses are entered by the subject himself/herself using the keyboard. In this paper we present two experiments regarding the development of such a procedure. The first experiment was set up in order to evaluate the effect of scoring just the keywords of each sentence instead of scoring each sentence as a whole. In the second experiment, two approaches of collecting and verifying keyboard responses were evaluated.

The nature of scoring (scoring whole sentences as a block, scoring all words separately, scoring only keywords, etc.) is known to have a range of effects on SRTn measurements. A comparison of speech reception tests indicates that scoring all words separately, instead of scoring the sentence as a whole, results in increased measurement efficiency, often enabling quicker measurements (Brand and Kollmeier, 2002). This is in accordance with the view of Hagerman and Kinnefors (1995) that fewer scoring units (whole sentences versus words) probably cause a higher standard deviation of individual SRTn values in repeated measurements. On the other hand, the sentence material of the Plomp and Mimpen test was not originally intended to be used with keyword or separate word scores. The test is based on balanced lists for optimal whole sentence scoring. Versfeld *et al.* (2000) reported a decrease in the slope of the discrimination function of 1.3%/dB when using the number of correctly repeated words instead of whole sentence scoring. In view of our objective to create an automatic procedure, we anticipated that a large set of keywords would provide the best basis. In that way, as much sentence information as possible is used, combined with a sufficient number of scoring units.

Apart from the use of keyword scoring, the adaptive procedure for reaching the final SRTn was also investigated in the first experiment. That is, we evaluated different strategies for the scoring of keywords and for the step size of the SNR adaption (viz., increasing or decreasing the SNR of the next sentence). The Plomp and Mimpen test uses an open set of sentences. As the number of words and keywords per sentence varies, scoring (key) words separately becomes more complicated with an open set of sentences. However, relatively few of the open-set tests score whole sentences as a block. In fact only the tests by Plomp and Mimpen (1979), Gelfland *et al.* (1988), and Nilsson *et al.* (1994) do. Other open-set SRTn-tests, such as the speech perception in noise (SPIN)-tests (Kalikow *et al.*, 1977; Bilger *et al.*, 1984) and the Göttingen sentence test (Kollmeier and Wesselkamp, 1997), calculate scores in percentage correct or only score a sentence correct if all *keywords* are repeated correctly [e.g., the open-set test by Macleod and Summerfield (1990)]. In our experiment we compared scoring per keyword (relative scoring) with scoring the entire set of keywords (block scoring).

On the basis of the scoring result, the step size of the SNR adaptation is determined when proceeding through the list of sentences. Most tests that score sentences as a block use fixed adaptation steps of ±2 dB [e.g., see Plomp and Mimpen (1979), Macleod and Summerfield (1990), and Nilsson *et al.* (1994)]. When (key)words are scored separately, more complicated stepping schemes can be developed. For example, the method used by Hagerman and Kinnefors (1995) consists of a fixed stepping scheme ranging from −2 to +3 dB in 1 dB steps, depending on the number of words repeated correctly. In our experiment we used a variant of this fixed step size scheme and also evaluated a more complex paradigm as proposed by Brand and Kollmeier (2002), where the step size effectively decreases as the sentence number in a list increases (gradual convergence toward the SRTn).

The second experiment focused on collecting responses using a computer keyboard (instead of verbal responses) and the problem of dealing with spelling and typing errors. A SRTn-test is supposed to measure speech reception and not the subject's spelling or typing abilities. How can one know whether something is a spelling or typing error and not a result of misperception? The obvious solution was to find a balance in which the errors that are likely to be spelling or typing errors are discarded, while errors that are probably due to misperception are penalized. By using a keyword approach from the first experiment, the influence of typing errors on the test results was reduced and the performance of the recognition algorithm was maximized. Two algorithms were compared: one simple spelling checker and a more complex dynamic alignment paradigm (Wagner and Fisher, 1974).

The effectiveness or success of the final automated procedure and of the intermediate results was determined by comparison with the standard Plomp and Mimpen test. This was done in terms of the estimated SRTn value, the test-retest reliability, and the shape/slope of the discrimination function. Before describing the two experiments, the selection of keywords and the adjustments to the adaptive procedures are discussed in the following section.

II. ADAPTIVE PROCEDURES

A. Selection of keywords

A first step in automating the SRTn procedure was to make use of keyword scoring. When deciding what set of keywords should be used, two approaches were possible. The first approach (*bottom-up*) considered the words that should be included. This set started with all content words, the "true" keywords (e.g., nouns and adjectives), and was enlarged by all function words that clearly had an important semantic function within the sentence context. The second approach (*top-down*) started with the whole sentence and checked which words could be left out. These were usually articles and (personal) pronouns with a less important semantic function. In practice the application of the two approaches led to virtually the same result, corresponding with the extraction of "newspaper headlines" from the sentences. The number of keywords ranges from 3 to 5, with an average of 3.7 keywords per sentence.[1]

B. Scoring and step size

With a limited number of keywords per sentence, the decision of a correct score can be made in two ways: either all keywords must be correct (block scoring) or scores are given for each keyword that is reproduced correctly (relative scoring). Referring to Hagerman and Kinnefors (1995), we

TABLE I. Stepping scheme for keyword scoring for sentences containing three to five keywords (e.g., a score of 2/5 means two out of five keywords correct).

Score	0/5	1/5	1/4	1/3	2/5	2/4	3/5	2/3	3/4	4/5	5/5
Step (dB)	+2	+2	+1.5	+1.5	+1	0	−1	−1.5	−1.5	−2	−2

considered a fixed stepping scheme based on the number of keywords that are repeated correctly. We developed a symmetrical stepping scheme for measuring the 50% threshold (Table I). A complicating factor was the fact that in the Plomp and Mimpen test the number of keywords per sentence is not fixed but varies from 3 to 5. With a 1.5 dB step size for the 1/4, 1/3, 2/3, and 3/4 scores, the five step scheme shows a balanced distribution of step sizes.[2] Note that in the block scoring condition this scheme results in a fixed step size of ±2 dB since in this condition the score is either 0 or 1.

Brand and Kollmeier (2002) developed a generalization of the Hagerman and Kinnefors method in which the change in the presentation level of the subsequent sentence is calculated on the basis of the following variables: *percentage of correct (key)words of the previous sentence (prev), number of reversals of the presentation level (i)*, and constant *target discrimination value (tar)*, and *slope of the discrimination function (slope)*. In formula,

$$\Delta L = -\frac{f(i)(\text{prev-tar})}{\text{slope}}. \quad (1)$$

The function $f(i)$ returns a smaller number as i increases; as a result the step size decreases as the list proceeds. This procedure is used either with an open or a closed set of sentences [the Göttingen and the Oldenburger sentences, see Kollmeier and Wesselkamp (1997) and Wagener *et al.* (1999a, 1999b, 1999c)] and could be implemented for the Plomp and Mimpen test simply by filling in the constants. Tar was set to 0.5 because the Plomp and Mimpen test measures the 50% speech reception threshold. The slope parameter was set to 0.15 dB^{-1}, in accordance with the slope of the discrimination function found by Plomp and Mimpen (1979).

In summary, two different scoring strategies were investigated: *block scoring* (all keywords or nothing) and *relative scoring* (number of correct keywords re all keywords). Furthermore, two different step size strategies were investigated: *fixed stepping scheme* (Hagerman and Kinnefors style) and *dependent step size* (Brand and Kollmeier style).

III. EXPERIMENT 1: COMPARISON OF ADAPTIVE PROCEDURES

A. Methodology and materials

1. Stimuli and design

The original test material as developed by Plomp and Mimpen (1979) was used, consisting of ten lists of 13 sentences spoken by a female speaker. All sentences were available in digital form with a sampling rate of 44,100 Hz and a 16 bit resolution. The keyword strategy was evaluated in four conditions (two scoring conditions × two step size conditions) against the standard procedure, which makes a total of five test conditions.

2. Subjects

A total of 15 subjects with normal hearing participated in the listening experiment. Normal hearing was defined as having a pure-tone threshold not exceeding 15 dBHL at any frequency from 250 to 2000 Hz and not exceeding 25 dBHL in the range of 2000–8000 Hz for both ears. The better ear was used for testing. The subjects might have experience with speech-in-noise testing, but none of the subjects were familiar with the Plomp and Mimpen sentence material.

3. Procedure

The ten lists were presented in a fixed order. Each of five conditions was tested with two lists. The conditions were tested in alternate orders, according to a 5×5 latin-square design. Having 15 subjects, each sequence was presented to 3 subjects. Three lists of sentences from the female speaker of Versfeld *et al.* (2000) were presented as practice materials.

The signal was delivered monaurally by headphones (Sony MDR-7509) through a RME Hammerfall DSP Multiface sound module. The noise level was fixed at a comfortable level of 70 dBA. The mixing of the speech and noise signal was done by a computer program according to the adaptive procedure under test. The subjects received written instructions. The subjects' task was to repeat the sentences verbally. As in the original test, each first sentence of a list was presented at a SNR of −8 dB. This sentence was repeated, each time at a 4 dB higher level, until the subject could reproduce it correctly (in the case of the keyword conditions, i.e., all keywords correct). The remaining 12 sentences were presented once in an up-down procedure according to the strategy under test. The tests were administered by an experimenter, who evaluated the responses.

4. Data analysis

The test-retest reliability was defined in accordance with Plomp and Mimpen (1979), viz., as the root mean square of the differences between the two SRTn values for each subject in each condition, divided by $\sqrt{2}$ [see also Nilsson *et al.* (1994), Hagerman and Kinnefors (1995), Versfeld *et al.* (2000), Brand and Kollmeier (2002)]. A permutation test was performed to check whether the differences among conditions were significant. For each condition 20 permutations of 10 out of 15 were drawn from the original data. (The number 10 was chosen because this is the number of subjects used by Plomp and Mimpen (1979) in their original evaluation.) For each of these samples, the test-retest reliability was estimated. Subsequently, an analysis of variance was performed on the 5×20 resulting values.

The discrimination function results from plotting the proportion of correct responses over all subjects for all presentation levels, thus indicating the intelligibility score as a function of presentation level. Because the number of sentences per individual subject was too small for a slope estimate, all data were pooled across subjects [see Plomp and

TABLE II. Mean SRTn (standard deviation in parentheses), test-retest reliability, and slope of the discrimination function of SRTn measurements for the five test conditions of experiment 1. Mean SRTn refers to the value across presentation levels; ML SRTn refers to the value obtained with the ML estimation.

Condition	Mean SRTn (SD) (dB)	Test-retest reliability (dB)	Slope (%/dB)	ML SRTn (dB)
0: standard test	−5.63 (0.86)	0.75	21	−5.65
1: relative/fixed	−6.99 (0.87)	0.69	15	−6.76
2: block/fixed	−5.83 (1.16)	1.17	16	−5.76
3: relative/dependent	−7.12 (0.86)	0.86	16	−7.13
4: block/dependent	−5.61 (1.45)	1.54	14	−5.39

Mimpen (1979)]. Note that this required that all subjects behaved in the same manner. In the block scoring conditions, the proportion simply equaled the relative number of sentences reproduced correctly. In the relative scoring conditions, the proportion of correct responses was obtained by averaging the relative number of keywords that was reproduced correctly for each sentence presented at each presentation level. The function was fitted by the maximum-likelihood (ML) criterion as developed by Versfeld *et al.* (2000), in which the SRTn and the slope are estimated in an iterative procedure. Besides a slope value, the ML estimation of the discrimination function also produces an estimate of the SRTn.

B. Results

Mean SRTn values in the five conditions are given in Table II. The mean values are based on the average of 30 individual SRTns per condition, as each subject was tested with two lists in each condition. The results show that the block scoring conditions (2 and 4) produce virtually the same SRTn as the standard test. The relative scoring conditions (1 and 3), on the other hand, seem to lead to lower SRTn values in comparison with the standard test. A repeated-measure analysis of variance indicates a significant difference among conditions ($p<0.01$). A subsequent Tukey-HSD *post hoc* test shows that both relative scoring conditions differ significantly ($p<0.01$) from the standard test, which indicates an effect of the scoring method on the SRTn value. An analysis of variance with *scoring* and *step size* as fixed factors (discarding data of the standard test) shows a significant effect of scoring ($p<0.01$), no effect of step size, and no interaction between factors.

To see if a learning effect occurred, the first and second measurements were compared for each subject in each condition. The results indicated no significant difference averaged over all conditions. To check whether any of the effects were individually significant, separate paired *t*-tests were performed per condition. No learning effect was observed.

In terms of test-retest reliability, the results in the relative scoring conditions are comparable to those of the standard test (values of 0.69 and 0.86 dB against 0.75 dB, Table II). With block scoring, the figures are considerably higher (1.17 and 1.54 dB). A repeated-measure analysis of variance

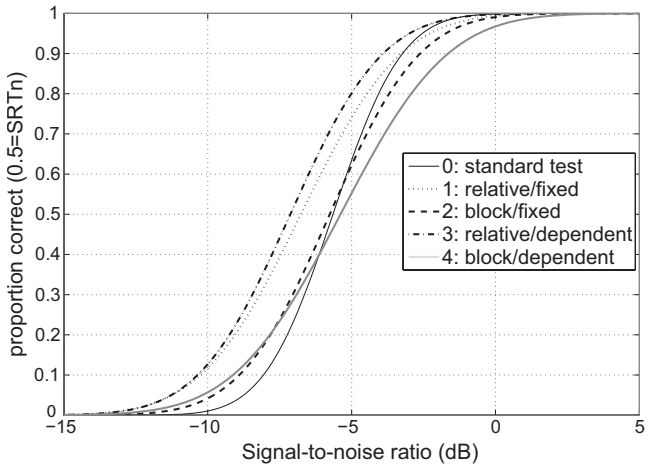


FIG. 1. Discrimination functions for the measurement of the SRTn in the five test conditions in experiment 1. See text and Table II for more details on slope values.

on the permutation resamples indicated a significant difference among conditions ($p<0.01$). A subsequent Tukey-HSD *post hoc* test, however, showed that the relative scoring conditions (1 and 3) did not differ significantly from the standard test.

To test whether pooling the data across subjects was justified for the estimation of the discrimination functions, an analysis of variance was conducted on the standard deviation within lists (i.e., the standard deviation of the mean score over the last ten sentences per list) with subject and condition as fixed factors. The results showed a significant effect of condition but no effect of subject and no interaction between subject and condition, thus indicating that all subjects behaved in the same manner. To see if a learning effect occurred, the first and second measurements were compared for each condition in a series of paired *t*-tests. No learning effect was observed. Therefore, all data were pooled across subjects for the estimation of the discrimination functions. Discrimination functions are shown in Fig. 1. Slope values and accompanying estimated SRTns (ML SRTn) are presented in Table II. Slope values range from 14%/dB (condition 4: relative/dependent) to 21%/dB (condition 0: standard test).

C. Discussion

1. SRTn

Reported common values (average SRTns for subjects with normal hearing) for the standard test (all words correct) range between −4.5 and −5.8 dB, depending on the sentence material [see Smits *et al.* (2004) for an overview]. The mean monaural SRTns found by Plomp and Mimpen (1979) are −5.6 dB (left ear only) and −6.2 dB (right ear only).

The results show that blocked keyword scoring produces the same SRTn as the standard test, whereas relative keyword scoring leads to a lower SRTn. This difference has an easy explanation: in block scoring any error directly leads to a zero score, where relative scoring often leads to a positive score. The effects on the progression of the adaptive procedure can be large. For example, three out of four keywords reproduced correctly yields a zero score in block scoring,

leading to an *increase* in the SNR for the next sentence, against a score of 0.75 in relative scoring, which is above the 50% threshold and leads to a *decrease* in SNR for the subsequent sentence. Hence, a block scoring threshold is different from a relative scoring threshold. For the sake of comparing the SRTn values with the standard test, however, SRTns acquired with the relative scoring procedures could be transformed by adding 1.4 and 1.5 dB for fixed and dependent step sizes, respectively.

As mentioned in Sec. III A 4, the ML estimation of the discrimination function, besides a slope value, also produces an estimation of the SRTn. The ML SRTn agrees well with the mean of the SRTns that were estimated by averaging across presentation levels, with a maximum difference of 0.23 dB (Table II).

2. Test-retest reliability

In terms of the standard deviation of individual SRTn values, the relative scoring conditions show results similar to the standard test. Among the keyword conditions, the highest test-retest reliability was found for the relative scoring condition in combination with a fixed stepping scheme (0.69 dB). Note that the figure we found for the standard test (0.75 dB) is lower than the standard deviation of 0.9 dB found by Plomp and Mimpen (1979). A closer look at the differences among the test conditions and the influences of the scoring factor indicates that relative scoring produces a higher test-retest reliability than block scoring. This confirms the prediction based on Hagerman and Kinnefors (1995) that scoring all keywords separately has a positive effect on the test reliability.

Regarding step size, however, the results are not in accordance with the original expectations. The data show a trend that the dependent step size conditions have a lower test-retest reliability than their fixed step size equivalents (Table II). The results, however, are in accordance with experience gained in practice. During the testing sessions, the adaptation of the presentation level in the dependent step size conditions repeatedly seemed to become too small, which resulted in an improbably high or low SRTn value for the given subject (based on the preceding and following measurements). The point is that although the dependent step size adaptive method reduces the standard deviation within measurements (i.e., the standard deviation of the presentation levels in single SRTn-measurements), it seems to produce less consistent SRTn estimates than the fixed step size method.

3. Discrimination function

The discrimination functions (Fig. 1) show that the standard test has the steepest slope of all test conditions. The slope value found for the standard test is 21%/dB, which is high, not only compared to the slope of 15%/dB found by Plomp and Mimpen for the exact same test,[3] but also compared to the slope values found in the literature for other tests [e.g., see Bilger *et al.* (1984), Gelfland *et al.* (1988), Nilsson *et al.* (1994), Kollmeier and Wesselkamp (1997), and Wagener *et al.* (2003)]. Among the other test conditions, the results show no large differences, slope values ranging from 14%/dB to 16%/dB.

For all experimental keyword conditions, the slope of the discrimination function is shallower than the slope that we found for the standard test. This may be due to the normalization of the sentence material. The rms levels of the sentences have been adapted so that each sentence became equally intelligible. Verifying only the keywords could interfere with the normalization because the rms of the keywords could differ from the rms of the whole sentence, and the size of the difference could vary among sentences. Consequently, if verified on keywords, sentences are no longer necessarily equally intelligible, which could have a negative effect on the test's discrimination function. The slope of the discrimination function of the tested methods might be improved by renormalizing the sentence material, which can be done by calculating the difference between the rms of the whole sentence and the rms of only the keywords for each sentence and by correcting the normalized sentence levels for the differences.

D. Summarizing remarks

The combination of relative scoring and a fixed stepping scheme produced the results that are the most similar to the standard test (condition 1 in Table II). The test-retest reliability is equivalent, but the slope of the discrimination function is shallower than the slope that we found for the standard test. However, our slope is comparable to other tests (Gelfland *et al.*, 1988; Wagener, 1999a, 1999b, 1999c; Versfeld *et al.*, 2000; Wagener *et al.*, 2003). Therefore, we used the combination of relative scoring with a fixed stepping scheme for further development of the automatic procedure.

IV. THE AUTOMATIC VERIFICATION OF RESPONSES

A. Verification procedures

The next step in automation is the verification of responses. We have explored two approaches to deal with this issue. The first approach is to use a spelling checker that consists of two steps: First, the spelling and typing errors are corrected; subsequently the responses are matched to the keywords in question. The outcomes of this approach depend on the dictionary and rules of the spelling checker. The second approach combines the aspects of correction and matching in one step of determining string similarity by dynamic alignment. Dynamic alignment, in its classical approach, is based on the minimal edit distance between the input string and a reference string (Wagner and Fisher, 1974; Yang *et al.*, 2003). For example, the deviation of the word *task* in reference to the word *token* can be expressed as a substitution (*a* for *o*), an insertion (*s*), and two deletions (*e* and *n*). In calculating the minimal edit distance, different deviations can be treated differently, so the effect of insertions, deletions, and substitutions can be individually regulated, and the likelihood of different substitutions can also be taken into account.

A more detailed description of the two methods is given in Sec. IV C. The first step in the development of verification

algorithms was to establish what kinds of errors occur when subjects have to enter their responses by keyboard. In order to do so, a pilot experiment was set up, which is presented below.

B. Pilot experiment: Error analysis

1. Methodology and materials

A total of seven subjects with no reported hearing problems and with varying computer experience and typing skills were asked to take an automatic test. Ages varied from 20 to 60 years. Each subject took the test repeatedly (varying from two to ten times), each time using different sentences. The lists were presented in a fixed order using the adaptive SRTn procedure that was selected from experiment 1 (relative keyword scoring with a fixed stepping scheme).

The subjects received written instructions. Their task was to reproduce the sentences by entering them by keyboard in the edit window of the computer program. The sound signal was delivered binaurally by headphones (Philips SBC HP140).

The verification of the responses was done by a simple *one-to-one* string matching algorithm. According to the relative scoring/fixed stepping scheme adaptive procedure as presented in Sec. III, each keyword was scored separately, and the presentation level of the next sentence was adapted dependent on the relative number of keywords that were reproduced correctly. For the remaining part, the test procedure was the same as described in Sec. III.

2. Results

A total of 416 responses were recorded, containing 1876 words. Fifteen of these words contained a typing or spelling error, constituting an error rate of 0.8%. Twelve were typing errors and three were spelling errors. Furthermore, 12 of the 15 errors appeared in a keyword, which means that on average each measurement contained 0.4 false negatives (keyword scored false due to a typing or spelling error). In ten cases the response was a correctly spelled Dutch word, strongly resembling the intended keyword. Examples hereof are "We *konden* nog niet in het gebouw" for "*Honden* mogen niet in het gebouw" and "Er wordt hier geen *lijst* opgebouwd" for "Er wordt in dit land geen *rijst* verbouwd." Following the strict criterion as used by Plomp and Mimpen (1979), these words were counted as an erroneous response and not as a typing or spelling error.

3. Discussion and conclusions

Given the low error rate, it may seem questionable whether it is necessary to have an algorithm that deals with typing or spelling errors. If the error rate was 2.5 times as high, so that each measurement contained one false negative, the effect on the final value of the SRTn could be at most 0.3 dB.[4] However, it is important to rule out the theoretical possibility that the test outcomes can be dependent on the listeners' typing and spelling skills. In other words, the low error rate does not rule out the need for an algorithm.

The errors mainly consisted of typing errors. Furthermore the results confirm that the implementation of the algorithms (choosing and evaluating the parameter values) should focus on the differentiation between typing and spelling errors and true errors that strongly resemble a keyword.

C. Implementation

1. Spelling checker

The spelling checker algorithm that was used is a very straightforward string matching algorithm that consists of three steps. The response is converted into a list of words, and each item on the list is checked in order to determine whether it is a known word in the dictionary. (It was verified that all words marked as keywords in the Plomp and Mimpen sentences were in the dictionary.) If not, the word is replaced by the first variant that is suggested by the spelling checker (e.g., the response *ryst* is not in de dictionary and is replaced by the first suggestion *rist*). Subsequently, each of the reference keywords is matched against the response list by one-to-one string matching. If a keyword occurs in the list, the matching function returns true and the score is incremented. Only one occurrence of a keyword is scored.

2. Dynamic alignment

The basic routine of the dynamic alignment algorithm is based on an algorithm that is used to compare, e.g., protein or DNA sequences. The algorithm first creates a two-dimensional matrix in which the two strings to be matched (the reference keyword and the response word) are set out on the x-axis and y-axis. Subsequently, the matrix is filled by comparing each character on one axis with all of the characters on the other axis. Dependent on the outcomes of the comparison, a score or penalty is obtained: 10 for a match, −4 for a deletion or insertion (gap), −4 for a substitution (mismatch) that is within one distance on the keyboard, and −7 for all other substitutions. In order to reduce the influence of mistakes in the sequence of characters within a word (e.g., *faeture* for *feature*), the first two substitutions with a character that is in the reference keyword are not penalized but considered a match. After the matrix is completely filled, the optimal alignment is determined by tracing back the shortest path across the matrix and summing the scores of the individual steps. A matching index is obtained by dividing the alignment score by the maximal possible score (which is the length of the reference keyword times ten). For example, the best path of the word *task* in reference to the word *token* is 10 (match *t*), −7 (substitute *a* for *o*), −4 (insert *s*), 10 (match *k*), −4 (delete *e*), and −4 (delete *n*), which results in an alignment score of 1 (10−7−4+10−4−4) and subsequently a matching index of 0.02 (1/50). The matching index is consequently compared with a threshold value, the matching criterion. The matching criterion was determined by testing the performance on the words found in Sec. IV B. With an equal error rate[5] of 23%, the optimum matching criterion was found to be 0.767. With a matching index above this value, the responded word is considered to be a match; otherwise, it is a mismatch. In this algorithm, each of the keywords is aligned with each item on the list of responded words in the same manner as described for the spelling checker algorithm.

V. EXPERIMENT 2: EVALUATION OF THE AUTOMATIC PROCEDURE

A. Methodology and materials

1. Stimuli and design

The original set of sentences (Plomp and Mimpen, 1979) was used for the test. The two automatic verification procedures (keyboard responses with either spelling checker or dynamic alignment paradigm) were evaluated against the original procedure of verbal responses and verification by an experimenter. This experiment had two goals: the practical goal of validating the automatic tests against the test procedure with verbal responses and the goal of testing the performance of the two verification algorithms themselves.

2. Subjects

A total of 12 subjects with no reported hearing problems participated in the experiment. Six of them were young people (ages ranging from 21 to 23) who were experienced in using the keyboard, and six of them were elderly people (ages ranging from 63 to 67) who were less experienced with a keyboard. For each subject, the pure-tone thresholds in both ears were tested. The better ear was used for testing. Some subjects were familiar with speech-in-noise testing, but none of the subjects were familiar with the Plomp and Mimpen sentences.

3. Procedure

The ten lists were presented in a fixed order. Each of three conditions was tested with three lists. The conditions were tested in alternate order, according to a 3×3 latin-square design. The first list was presented as a practice list.

The test was administered by a computer program. The sound signal was delivered monaurally by headphones (Sony MDR-7509) through a SBLive! 400 sound module. The noise level was fixed at a comfortable level of 70 dBA. The subjects received written instructions. The subjects' task was to reproduce the sentences by entering their responses into the designated edit window using the keyboard. In the standard test condition, the responses were entered by an experimenter, the subjects' task being to repeat the sentences verbally.

The responses were evaluated with the verification algorithm under test. Each condition used the relative score/fixed stepping scheme adaptive procedure as described in Sec. IV B 1, with one small alteration. In the original procedure, the first sentence of a list was repeated until the listeners' response was entirely correct. To prevent the possibility of an infinite loop caused by a repeated misspelling of one of the keywords, the number of repetitions of each first sentence of a list was limited to 4.

4. Data analysis

The test procedures were evaluated on the estimated SRTn value, the test-retest reliability, and the slope of the discrimination function. We defined the test-retest reliability by the rms of the differences between two SRTn values for each subject, divided by $\sqrt{2}$ [see experiment 1 and Plomp and Mimpen (1979)] for a situation with two measurements per subject per condition. As the present test consisted of three measurements per condition per subject, we took the average of the three possible rms/$\sqrt{2}$ values as the value for test-retest reliability. This procedure enabled us to make a direct comparison with the results of experiment 1.

The spelling checker and dynamic alignment algorithms were also evaluated directly on the *precision*$\times$*recall* ratio. Precision is calculated as the number of hits divided by the sum of the numbers of hits and false alarms. Recall denotes the ratio of the number of hits to the total number of typing or spelling errors. Both ratios are expressed as a percentage.

TABLE III. Mean SRTn (standard deviation in parentheses) for the three test conditions of experiment 2, overall, and differentiated for the young and elderly groups.

	Mean SRTn (SD) (dB)		
Condition	Overall	Young	Elderly
0: verbal repetition, experimenter	−6.83 (1.09)	−7.17 (0.94)	−6.49 (1.15)
1: keyboard, dynamic alignment	−6.99 (1.17)	−7.18 (1.28)	−6.79 (1.03)
2: keyboard, spelling checker	−6.73 (0.98)	−7.01 (0.94)	−6.42 (0.96)

B. Results

1. SRTn

Mean SRTn values in the three test conditions are given in Table III, both overall and differentiated between the groups young and elderly. The overall mean values are based on the average of 36, 35, and 34 individual SRTns for the conditions verbal, keyboard/dynamic alignment, and keyboard/spelling checker, respectively. Each subject was tested with three lists in each condition; however, the data contain three outliers, all of which occurred in the elderly group. The three outlier values are discarded in the analysis and are described in Sec. V C.

Both automatic verification conditions produced virtually the same SRTns as the verbal reference test. A repeated-measure analysis of variance showed no difference among conditions. Mean SRTns of repeated measurements are presented in Table IV. No effect of learning or fatigue was observed.

TABLE IV. Means for the first, second, and third measurements of SRTns in experiment 2. Mean SRTn refers to the value averaged across presentation levels.

Condition	Mean SRTn 1 (dB)	Mean SRTn 2 (dB)	Mean SRTn 3 (dB)
0: verbal repetition/experimenter	−6.72	−7.02	−6.76
1: keyboard/dynamic alignment	−7.01	−6.95	−7.00
2: keyboard/spelling checker	−6.56	−6.90	−6.71
Overall	−6.77	−6.95	−6.82

TABLE V. Mean SRTn, test-retest reliability, and slope of the discrimination function for the three test conditions of experiment 2. Mean SRTn refers to the value averaged across presentation levels; ML SRTn refers to the value obtained with the ML estimation.

Condition	Mean SRTn (dB)	Mean test-retest reliability (dB)	Slope (%/dB)	ML SRTn (dB)
0: verbal repetition/experimenter	−6.83	0.65	14.8	−6.96
1: keyboard/dynamic alignment	−6.99	0.82	13.4	−7.18
2: keyboard/spelling checker	−6.73	0.75	12.5	−6.96

2. Test-retest reliability

In the analysis of the test reliability, the outlier data points of one list were replaced by the average of the values of the remaining two lists. In terms of test-retest reliability, values vary from 0.82 dB for the dynamic alignment condition to 0.65 dB for the reference test (Table V).

In order to check whether the differences were significant, a permutation test was performed. For each condition 20 permutation samples of 10 out of 12 were drawn from the original data, and for each of these samples the test-retest reliability was calculated. An analysis of variance on the samples indicated a significant effect of condition ($p < 0.01$). A subsequent Tukey-HSD *post hoc* test showed that all conditions differ significantly from each other ($p < 0.01$).

3. Discrimination function

An analysis of variance (with subject and condition as fixed factors) on the standard deviation within lists (i.e., the standard deviation of the mean score over the last ten sentences per list) showed no significant effects or interactions, thus indicating that all subjects behaved in the same manner. To see if a learning effect occurred, the first, second, and third measurements were compared for each condition in a series of paired *t*-tests. No learning effect was observed. Therefore, all data were pooled across subjects for the estimation of the discrimination functions.

Discrimination functions are shown in Fig. 2. Slope values and accompanying estimated SRTns are presented in Table V. Slope values range from 12.5%/dB (keyboard, spelling checker) to 14.8%/dB (verbal repetition, experimenter). Besides a slope value, the ML estimation of the discrimination function also yields an estimation of the SRTn. The ML SRTn is very similar to the mean of the SRTns that were estimated by averaging across presentation levels, differences being at most 0.23 dB (see Table V).

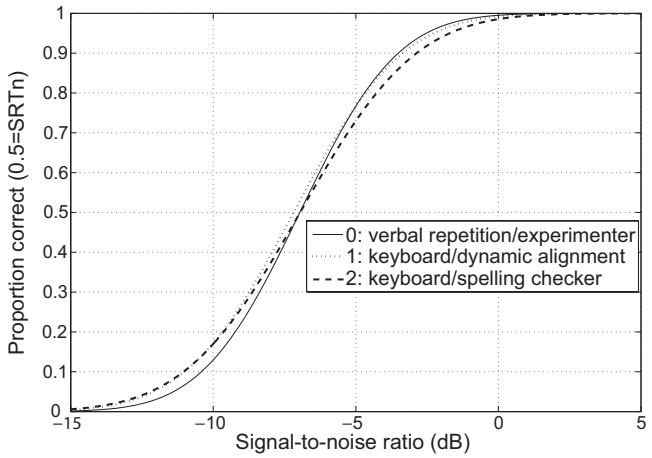


FIG. 2. Discrimination functions for the measurement of the SRTn in the three test conditions in experiment 2. See text and Table V for more details on slope values.

TABLE VI. Algorithm performance as derived from experiment 2.

Algorithm	Precision	Recall	Precision × Recall
1: dynamic alignment	62%	67%	42%
2: spelling checker	100%	29%	29%

4. Algorithm performance

In the keyboard conditions, a total number of 36 keywords contained a typing or spelling error. Fifteen errors occurred in the dynamic alignment condition, of which ten were positively matched. In six cases, a word that was spelled correctly in the context of the response, but was not equal to a keyword, was falsely labeled correct (false alarm). In the spelling checker condition, 21 errors were made, of which 6 were corrected. The spelling checker algorithm produced no false alarms. Precision × recall ratios are 42% and 29%, respectively, for the alignment and spelling checker algorithms (Table VI).

C. Discussion

1. SRTn

The results showed no significant differences in either mean SRTn or mean standard deviation within lists between the young and experienced and the elderly and less experienced keyboard users. Note that a very basic level of typing skill is assumed as people have to know how to type to be able to take the test. However, the results suggest no difference in test results when it comes to age group and self-reported typing skills.

2. Outliers

The three outliers all concern cases in which—due to repeated errors—the first sentence was presented the maximum number of four times before switching to the next sentence. As a result, the second sentence was presented at a SNR of +12 dB. In spite of a series of consecutive correct responses that resulted in presentation levels well below zero further into the procedure, the subjects were unable to reach a normal SRTn.

The outliers indicate that in the current procedure, the repetition of a typing or spelling error in the first sentence is not properly dealt with by the verification algorithm, which can result in a large underestimation of the SRTn. As a solution for this problem, the adaptive procedure was adjusted by allowing the first sentence to be only partially correct (i.e., one keyword is allowed to be incorrect) before switching to the second sentence, instead of requiring the first sentence to be completely correct.

3. Test-retest reliability and discrimination function

The results for the verbal repetition condition are an almost exact replication of the values found in Sec. III. In comparison to the verbal repetition task, both automatic verification conditions have slightly higher test-retest variability. Between keyboard conditions, the results are contradictory. On the test-retest reliability, the spelling checker algorithm performs better than dynamic alignment, whereas in the slope of the discrimination function the dynamic alignment algorithm performs better.

Values reported in the literature are roughly 1–1.5 dB for the test-retest reliability, and 10%–15%/dB for the slope of the discrimination function (group curves; data pooled across subjects). Compared to these values, both automated verification procedures score alike. In light of the reference data that we have found for the original test procedure, the test-retest reliability is maintained (0.65–0.75 dB for the reference, 0.75 or 0.82 for keyboard responses), but the slope has become less steep (21% per dB in experiment 1 to approximately 13% per dB for keyboard conditions in experiment 2). As suggested in Sec. III, the slope of the discrimination function of keyword verification might be improved by renormalizing the sentence material.

4. Algorithm performance

Results show that the dynamic alignment algorithm has the higher precision × recall ratio, although it should be noted that the impact of this observation is limited by the sparsity of errors. Whereas the spelling checker algorithm has a perfect precision, the recall is very low. The error pattern reveals two mechanisms underlying the high number of misses. Either the erroneous response is being replaced by the wrong alternative or the erroneous response is not replaced because it constitutes an existing word. To improve this requires far-reaching adjustments in the dictionary: basically the development of a task specific dictionary or even a set of dictionaries that are keyword specific. This would seriously limit the possibilities of the test being used with other sentence sets.

A closer look at the error pattern of the dynamic alignment algorithm shows an underlying round off problem that causes a lot of false mismatches. As it was implemented in the algorithm, the value of the matching criterion (0.767) is a round off of a fraction (23/30). However, the matching scores are obtained by dividing the score of the shortest path by the maximum score that is possible and are not rounded off. Hence, a word that scored 23/30 (or 46/60) was deemed a mismatch: a score that occurs, e.g., when a six-letter word contains one insertion or deletion. By slightly adjusting the criterion so that these words are scored as a match, the recall of the algorithm is given a boost to 93%. Although this also increases the number of false alarms, this does not affect the precision. As a result, the performance is altered to a precision × recall ratio of 60%.

VI. CONCLUSIONS

In the current paper, we have reported the development of a fully automated procedure for the measurement of the speech reception threshold for sentences in noise, which can be administered by computer. The procedure was based on the test developed by Plomp and Mimpen (1979), which consists of a repetition task of a simple straightforward sentence material that is presented against a background of stationary speech-shaped noise. The results of this study therefore hold for stationary noise and cannot be generalized for test procedures that use fluctuating noise.

The evaluation of different scoring strategies showed that adaptive procedures, which score each keyword separately (relative scoring), have a better test-retest reliability than procedures that score all keywords as a block. Furthermore, adaptive procedures that use a fixed step size show a higher test-retest reliability than procedures that use a converging step size (which decreases as the measurement proceeds). In comparison to the original procedure of whole sentence scoring, the slopes of the discrimination functions are less steep for all procedures that use keywords. However, slope values are still in line with data reported in the literature for other SRTn tests in stationary noise, with small differences between keyword conditions.

Only a limited number of typing and spelling errors were found when collecting responses from listeners using a keyboard. Based on this small data set, dealing with errors by means of a procedure based on a dynamic alignment resulted in less misses and less false alarms than when a simple spelling checker was used. Furthermore, the results indicated that switching from a verbal response task to manually entering the response by keyboard does not affect the test outcomes. In conclusion, the current study shows that it is feasible to obtain accurate SRTn data for sentences using an automated measurement procedure.

ACKNOWLEDGMENTS

This research was supported by grants from the European Union FP6, Project 004171 HEARCOM. The authors wish to thank Gerrit Bloothooft of Utrecht University for his contributions to earlier versions of the manuscript. Niek Versfeld of the Academic Medical Center in Amsterdam and three anonymous reviewers are acknowledged for their critical reading and useful comments.

[1]The entire set of 130 sentences, keywords marked, is available on request.

[2]On first appearance the 3 dB difference between the 1/3 and 2/3 correct scores may seem to be large. The reason behind this is that an adaptive procedure can only be accurate and discriminative if there are enough reversals during the track of the procedure. If a sentence contains only three keywords, the decision about the correctness is based on relatively little information. In such a case a relatively large step size is preferable over a small step size since it is more likely to force a reversal. In the case of 2/5 and 3/5 correct scores, the small difference (2 dB) is justified since it is based on relatively more information.

[3]It should be noted that the curve that was found by Plomp and Mimpen was fitted by eye, whereas we used a ML estimation. This methodological difference might—at least in part—explain the difference between their slope value (15%/dB) and the value that we found (21%/dB) for exactly the same test.

[4]Under the assumption that the up-and-down procedure is the same for the remaining part of the test, one keyword wrong can lead to a difference of at most 3 dB (i.e., an increase instead of a decrease of 1.5 dB) for the next sentence. As the SRTn is calculated by averaging ten presentation levels, this can result in a maximum increase of the SRTn of 0.3 dB.

[5]Equal error rate is generally defined as the error rate where the number of false mismatches and the number of false matches are approximately the same.

Bilger, R., Nuetzel, J., Rabinowitz, W., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32–48.

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**, 2801–2810.

Gelfland, S., Ross, L., and Miller, S. (**1988**). "Sentence reception in noise from one versus two sources: Effects of aging and hearing loss," J. Acoust. Soc. Am. **83**, 248–256.

Hagerman, B., and Kinnefors, C. (**1995**). "Efficient adaptive methods for measuring speech reception threshold in quiet and in noise," Scand. Audiol. **24**, 71–77.

Kalikow, D., Stevens, K., and Elliot, L. (**1977**). "Development of a test of speech intelligibility in noise using sentences with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Kollmeier, B., and Wesselkamp, M. (**1997**). "Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment," J. Acoust. Soc. Am. **102**, 2412–2421.

Letowski, T., Hergenreder, P., and Tang, H. (**1992**). "Relationships between speech recognition threshold, average hearing level, and speech importance noise detection threshold," J. Speech Hear. Res. **35**, 1131–1136.

Macleod, M., and Summerfield, A. Q. (**1990**). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise," Br. J. Audiol. **24**, 29–43.

Nilsson, M., Soli, S., and Sullivan, J. (**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.

Noordhoek, I., Houtgast, T., and Festen, J. (**1999**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners," J. Acoust. Soc. Am. **105**, 2895–2902.

Noordhoek, I., Houtgast, T., and Festen, J. (**2000**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners," J. Acoust. Soc. Am. **107**, 1685–1696.

Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," J. Speech Hear. Res. **29**, 146–154.

Plomp, R., and Mimpen, A. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.

Rooij, J. v., and Plomp, R. (**1991**). "The effect of linguistic entropy on speech perception in noise in young and elderly listeners," J. Acoust. Soc. Am. **90**, 2985–2991.

Smits, C., Kapteyn, T., and Houtgast, T. (**2004**). "Development and validation of an automatic speech-in-noise screening test by telephone," Int. J. Audiol. **43**, 15–28.

Smoorenburg, G. (**1992**). "Speech reception in quiet and in noisy conditions by individuals with noise-induced hearingloss in relation to their tone audiogram," J. Acoust. Soc. Am. **91**, 421–437.

Versfeld, N., Daalder, L., Festen, J., and Houtgast, T. (**2000**). "Method of selection of sentence materials for efficient measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.

Wagener, K., Brand, T., and Kollmeier, B. (**1999a**). "Entwicklung und Evaluation eines Satztests für die Deutsche Sprache I: Design des Oldenburger Satztests [Development and evaluation of a sentence test for the German language I: Design of the Oldenburg sentence test]," Z. Fuer Audiologie, Audiological Acoust. **38**, 4–15.

Wagener, K., Brand, T., and Kollmeier, B. (**1999b**). "Entwicklung und Evaluation eines Satztests für die Deutsche Sprache II: Optimierung des Oldenburger Satzetests [Development and evaluation of a sentence test for the German language II: Optimization of the Oldenburg sentence test]," Z. Fuer Audiologie, Audiological Acoust. **38**, 44–56.

Wagener, K., Brand, T., and Kollmeier, B. (**1999c**). "Entwicklung und Evaluation eines Satztests für die Deutsche Sprache III: Evaluation des Oldenburger Satztests [Development and evaluation of a sentence test for the German language III: Evaluation of the Oldenburg sentence test]," Z. Fuer Audiologie, Audiological Acoust. **38**, 86–95.

Wagener, K., Josvassen, J., and Ardenkjaer, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol. **42**, 10–17.

Wagner, R., and Fisher, M. (**1974**). "The string to string correction problem," J. Assoc. Comput. Mach. **21**, 168–173.

Yang, Q., Yuan, S., Zhao, L., Chun, L., and Peng, S. (**2003**). "Faster algorithm of string comparison," Pattern Anal. Appl. **6**, 122–133.

Identification of Mandarin tones by English-speaking musicians and nonmusicians

Chao-Yang Lee[a)]
School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio 45701

Tsun-Hui Hung[b)]
School of Music, Ohio University, Athens, Ohio 45701

(Received 21 May 2008; revised 29 August 2008; accepted 31 August 2008)

This study examined Mandarin tone identification by 36 English-speaking musicians and 36 nonmusicians and musical note identification by the musicians. In the Mandarin task, participants were given a brief tutorial on Mandarin tones and identified the tones of the syllable *sa* produced by 32 speakers. The stimuli included intact syllables and acoustically modified syllables with limited *F*0 information. Acoustic analyses showed considerable overlap in *F*0 range among the tones due to the presence of multiple speakers. Despite no prior experience with Mandarin, the musicians identified intact tones at 68% and silent-center tones at 54% correct, both exceeding chance (25%). The musicians also outperformed the nonmusicians, who identified intact tones at 44% and silent-center tones at 36% correct. These results indicate musical training facilitated lexical tone identification, although the facilitation varied as a function of tone and the type of acoustic input. In the music task, the musicians listened to synthesized musical notes of three timbres and identified the notes without a reference pitch. Average identification accuracy was at chance level even when multiple semitone errors were allowed. Since none of the musicians possessed absolute pitch, the role of absolute pitch in Mandarin tone identification remains inconclusive. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2990713]

PACS number(s): 43.71.Hw, 43.75.Cd [DD] Pages: 3235–3248

I. INTRODUCTION

Pitch perception is involved in both spoken language comprehension and music perception. For language, pitch is used for linguistic contrasts such as lexical tone and intonation (Ladefoged, 2003). For music, pitch is one of the most common dimensions for constructing an organized system of musical elements (Patel, 2008). Given the functional role of pitch in linguistic and musical contrasts, a legitimate question is whether the same perceptual mechanism is implicated in both linguistic and musical pitch perceptions.

This study explored the relationship between linguistic and musical pitch perceptions by examining identification of Mandarin tones and musical notes by English-speaking musicians and nonmusicians. In particular, we investigated how non-native listeners with or without a musical background dealt with identifying acoustically modified Mandarin tones produced by multiple speakers. The musicians also participated in a music task where they were asked to identify musical notes without a reference pitch. Our goal was to evaluate the nature of non-native lexical tone perception by incorporating common challenges to speech perception and to assess the role of absolute pitch in the musicians' advantage in lexical tone processing commonly reported in the literature (Alexander *et al.*, 2005; Gottfried, 2007; Gottfried and Riester, 2000; Gottfried *et al.*, 2001; Wong *et al.*, 2007).

Mandarin is a lexical tone language where pitch variation over a syllable is lexically contrastive. It has been established that *F*0 is the primary acoustic correlate of lexical tone (e.g., Tseng, 1981). Identification of musical notes certainly involves detecting pitch as well. In other words, identification of lexical tones and musical notes both involve mapping pitch information onto discrete linguistic or musical categories. If common perceptual mechanisms are involved, performance in musical and linguistic pitch perceptions should be correlated.

There is some experimental evidence on the relationship between music and lexical tone perception. Studies on lexical tone perception by non-native listeners showed that listeners with a musical background tend to identify lexical tones better than those without a musical background. Gottfried and Riester (2000), also reported in Gottfried, 2007, found that English-speaking listeners who were music majors identified Mandarin tones better than nonmajors. In addition, tone identification accuracy also correlated positively with performance in detecting the direction of nonlinguistic sine-wave glides. Gottfried *et al.* (2001), also reported in Gottfried, 2007, showed that English-speaking musicians were more accurate than nonmusicians in judging whether two stimuli had the same or different Mandarin tones. The musicians also performed better in vocally imitating the tones. Alexander *et al.* (2005) similarly found that English-speaking musicians were faster and more accurate than nonmusicians in Mandarin tone discrimination and identification. Wong *et al.* (2007) in an electrophysiological study found more robust brainstem encoding of linguistic pitch in

a)Electronic mail: leec1@ohio.edu
b)Present address: Program in Cognitive Ethnomusicology, School of Music, The Ohio State University, Columbus, Ohio 43210.

English-speaking musicians, who also showed better identification and discrimination of Mandarin tones in a behavioral task. Taken together, these results support the idea that extensive musical experience enhances the processing of linguistic pitch, suggesting an overlap of pitch processing mechanism in the two domains. Gottfried (2007), Gottfried and Riester (2000), and Wong *et al.* (2007) also hinted that the ability to track pitch direction and movement is implicated in the musicians' superior performance in lexical tone processing.

Recently, Deutsch and co-workers suggested that lexical tone is associated with absolute pitch, the ability to name or produce the note of a particular musical pitch in the absence of a reference note (Deutsch *et al.*, 2004a; Deutsch *et al.*, 2006). In particular, Deutsch and co-workers noted that individuals with absolute pitch are able to associate a particular pitch with a verbal label, just as tone language speakers associate a particular pitch or a combination of pitches with a linguistic tonal category. Therefore, tone language speakers could be said to possess a form of absolute pitch. Since absolute pitch is treated by tone language speakers as a feature of speech, tone language speakers would be expected to show evidence of absolute pitch.

To test this hypothesis, Deutsch *et al.* (2004a) asked speakers of Mandarin, Vietnamese (both tone languages), and English (a nontone language) to read a word list of their native language in several recording sessions across two days. Acoustic analyses of the recordings showed that the average *F*0 of the enunciations across recording sessions was more consistent for the tone language speakers compared to the English speakers. This result was interpreted as demonstration of a precise and stable form of absolute pitch in tone language speakers. Burnham *et al.* (2004) replicated this finding with a variant of the word production task. They noted, however, that the difference between the tone languages (3/4 semitone) and the nontone language (1 semitone) was not particularly substantial.

Additional evidence for the connection between absolute pitch and tone language came from an investigation on the prevalence of absolute pitch in musicians speaking a tone versus a nontone language. Deutsch *et al.* (2006) reported a significantly higher percentage of absolute pitch in musicians who speak Mandarin as their native language compared to those who speak English. In the absolute pitch task, participants listened to 36 piano notes and were asked to indicate the name of each note without a reference pitch. In general, absolute pitch was found to be more prevalent for people who started musical training at an earlier age. When the age of onset of musical training was taken into consideration, absolute pitch was substantially more prevalent in the Mandarin-speaking musicians, suggesting that the linguistic use of pitch may play a role in processing musical pitch categories.

If absolute pitch is indeed implicated in linguistic pitch processing, one would expect superior performance by individuals with absolute pitch in lexical tone processing. Although Deutsch *et al.* (2004a) and Burnham *et al.* (2004) demonstrated greater consistency of pitch production in tone language speakers, it is not clear whether the acoustic data could address how absolute pitch was manifested perceptually. While the prevalence of absolute pitch in Mandarin-speaking musicians (Deutsch *et al.*, 2006) suggests the predisposition of tone language speakers in acquiring absolute pitch, it is not clear whether absolute pitch actually contributes to lexical tone perception. Although previous research has consistently shown the advantage of musical background in lexical tone processing (Gottfried, 2007; Gottfried and Riester, 2000; Gottfried *et al.*, 2001; Alexander *et al.*, 2005; Wong *et al.*, 2007), Gottfried and Riester (2000), also reported in Gottfried, 2007, remains to our knowledge the only study that evaluated both linguistic and nonlinguistic pitch perceptions in the same group of individuals. To further explicate the nature of the association between linguistic and musical pitch perception and to test the absolute pitch hypothesis, it would be informative to assess performance in both lexical tone and musical pitch identification and to evaluate the correlation between the two tasks.

With these considerations, the present study included two perception experiments, one on lexical tone identification and the other on musical note identification. The Mandarin tone experiment evaluated how English-speaking musicians and nonmusicians dealt with learning to identify Mandarin tonal contrasts in the face of minimal experience with Mandarin, incomplete acoustic input, and speaker variability. In particular, speech signals are often incomplete and variable, yet human listeners are remarkably adept at uncovering linguistic representations intended by speakers. In the speech perception literature, the use of acoustically degraded stimuli has revealed important information about how human listeners handle various sources of acoustic variability to achieve perceptual constancy (e.g., Strange *et al.*, 1983). Dealing with speaker variability is also an integral part of speech processing (Johnson, 2005). The nature of the musicians' advantage in lexical tone processing may be revealed when these common challenges to speech perception are incorporated into test materials. The musical note experiment was a variant of the absolute pitch task used in Deutsch *et al.* (2006), assessing musicians' ability to identify musical notes without a reference pitch. With measures of both lexical tone and musical note identification, the association between Mandarin tone identification and absolute pitch could be evaluated.

In sum, the present study aimed to explore the relationship between linguistic and musical pitch processing by addressing the following questions: (1) What is the role of musical background in the perceptual processing of intact and incomplete Mandarin tones? If musical pitch processing abilities could facilitate linguistic pitch perception, musicians would be expected to perform better than nonmusicians in the Mandarin tone identification task. (2) What is the nature of the musicians' advantage in identifying Mandarin tones, if there is indeed an advantage? If absolute pitch is implicated (Deutsch *et al.*, 2004a; Deutsch *et al.*, 2006), there should be a positive correlation between performance in the Mandarin tone identification task and in the musical note identification task.

II. EXPERIMENT 1: MANDARIN TONE IDENTIFICATION

The stimuli in the Mandarin tone identification task included intact and acoustically modified Mandarin syllables produced by multiple speakers. The modified syllables included "silent-center" and "onset-only" syllables, where the majority of the voiced portion of a syllable was attenuated to silence and devoid of *F*0 information. As noted, the purpose of introducing acoustically degraded stimuli and speaker variability was to create a task that would incorporate common challenges to speech perception. It has been shown that silent-center Mandarin tones could be identified quite accurately by native and non-native listeners despite missing substantial *F*0 information, indicating that listeners are capable of reconstructing lexical tones based on limited acoustic input (Gottfried and Suiter, 1997; Lee *et al.*, 2006, 2008a). Furthermore, incomplete stimuli also allow examination of specific pitch dimensions (e.g., pitch direction and pitch height) that may contribute to tone identification.

Dealing with speaker variability is also an essential part of lexical tone processing. Given that *F*0 range varies across individuals, the absolute *F*0 of a given tone is likely to differ across speakers; different tones are also likely to show *F*0 overlap across speakers. A number of studies have shown that some speaker normalization process is implicated in lexical tone perception (Leather, 1983; Lin and Wang, 1984; Moore and Jongman, 1997; Wong and Diehl, 2003; Zhou *et al.*, 2008) and that speaker variability imposes a processing cost on the identification of intact and incomplete Mandarin tones (Lee *et al.*, 2008b). In the present study, the stimuli included syllables produced by 16 female and 16 male speakers. Acoustic analyses would be conducted to verify that the purported speaker variability actually exists.

A. Method

1. Materials

The Mandarin syllable *sa*, produced with all four tones by 16 female and 16 male native speakers, was selected to generate the stimuli for this experiment. The recordings were made in a sound-treated booth with an Audio-technica AT825 field recording microphone connected through a preamplifier and A/D converter (USBPre microphone interface) to a Windows personal computer. The speakers were instructed to read the syllables in citation form. The recordings were digitized with the Brown Laboratory Interactive Speech System (BLISS, Mertus, 2000) at 44.1 kHz with 16 bit quantization. Each syllable was identified from the BLISS waveform display, excised from the master file, and saved as an audio file. The peak amplitude was normalized across syllables with the "scale to maximum" function of BLISS.

Each *sa* syllable was digitally processed with BLISS to generate three types of syllables: intact, silent-center, and onset-only syllables. The silent-center syllables were generated by removing all but the first and final 15% of the voiced portion. The onset-only syllables were generated by removing all but the first 15% of the voiced portion. The fricative [s] was preserved in all intact and modified syllables. The removed parts were digitally "silenced" such that the overall duration remained the same as that of the intact syllables.

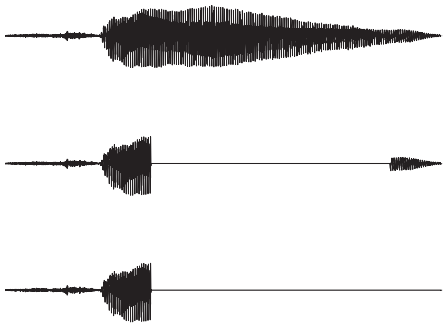


FIG. 1. Illustration of stimulus construction in the Mandarin tone experiment. From top down: Intact, silent-center, and onset-only syllables. Silent-center syllables were generated by digitally silencing the 70% of the middle of the voiced portion. Onset-only syllables were constructed by silencing all but 15% of the beginning of the voiced portion.

There were no perceptible clicks as a result of the signal processing; therefore, no further tapering procedure was applied. An example of the acoustic modifications is shown in Fig. 1. A total of 384 stimuli (4 tones $\times$ 32 speakers $\times$ 3 modifications) were used in the experiment.

2. Participants

Participants included 36 English-speaking musicians and 36 age- and gender-matched nonmusicians. They were recruited from the student population at Ohio University with cash compensation. All reported normal hearing, speech, and language. No one had prior experience with Mandarin or any tone languages.

The musicians included 20 females (mean age =24 years, SD=4.5) and 16 males (mean age=21 years, SD=2.6). All were junior, senior, or graduate music majors in the School of Music at Ohio University. Their area of concentration included performance, voice performance, music education, music therapy, music theory, composition, and history/literature. All musicians were required to demonstrate competency in a major and a secondary instrument among other musical proficiency requirements (Ohio University School of Music Undergraduate Handbook, 2007–2008). The average age at which the musicians first received their music training was 9.4 years (SD=3.2). The nonmusicians included 20 females (mean age=20 years, SD=1.36) and 16 males (mean age=21 years, SD=1.91). No one reported any formal musical training or substantial musical learning experience.

3. Procedure

The stimuli, saved as individual audio files, were imported to AVRunner, the subject-testing program in BLISS, for stimulus presentation and response data acquisition. The experiment was divided into three sections: the 128 intact syllables were presented first, followed by the 128 silent-center syllables, and finally the 128 onset-only syllables. Considering the challenging nature of the task (i.e., identifying a novel linguistic contrast in an unfamiliar language with minimal instructions), the presentation of the stimuli was intentionally ordered from those with the most complete

acoustic input to those with the least acoustic input. Breaks were given between sections, when instructions for the next section were given.

For each section, the 128 stimuli produced by the 32 speakers were assigned to four blocks such that each block included only one stimulus from a given speaker. In other words, each block had 32 stimuli and all stimuli were produced by different speakers. Within each block, the number of male and female speakers was balanced (16 male and 16 female), so was the number of the four tones (eight stimuli for each of the four tones). For each participant, AVRunner assigned a uniquely randomized presentation order such that no two participants received the same order of presentation. The order of presentation for the blocks was also randomized for each participant. A 10 s break was given between blocks.

Participants were tested individually in a sound-treated room. The experimenter first explained the lexically contrastive function of tones in Mandarin by giving minimal tone pairs as examples. The *F*0 contours of the four Mandarin tones were drawn on a blackboard to illustrate the differences among the four tones, which were also verbally described to the participants: the high-level tone (tone 1) begins with a high *F*0 and stays level throughout; the high-falling tone (tone 4) begins with a high *F*0 and moves downward to a low *F*0; the low-rising tone (tone 2) begins with a low *F*0 and moves upward to a high *F*0; and the low-dipping tone (tone 3) begins with a low *F*0 and either stays low or makes a dip in *F*0.

In accordance with these descriptions, the participants were directed to four keys on the keyboard labeled with "→," "↘," "↗," and "∨," representing the high-level (tone 1), high-falling (tone 4), low-rising (tone 2), and low-dipping (tone 3) tones, respectively. Despite the common use of number designations for Mandarin tones, the number system was never introduced to the participants to avoid any additional processing demand of memorizing the arbitrary association between *F*0 patterns and numbers. The four labels were placed on two rows, indicating the relative height of beginning *F*0 of the tones. In particular, keys representing the high-level (→) and high-falling (↘) tones were placed on the upper row and those representing the low-rising (↗) and low-dipping (∨) tones were placed on the lower row. The participants were told that both *F*0 shape and *F*0 height were involved in the identity of a tone; therefore, both should be considered when making tone judgments. The participants were also told that since the syllables were produced by multiple male and female speakers, the *F*0 range was likely to differ across speakers. Therefore, tone judgments should be based on the tones intended by specific speakers.

The participants listened to the stimuli through a pair of headphones connected to a Windows personal computer. Ten practice stimuli, none produced by any speakers in the actual experiment, were given prior to each section to familiarize the participants with the procedure and the stimulus/response format. The participants were instructed to identify the tone of each syllable by pressing one of the four labeled buttons. They were also told that their responses would be timed and that they should attempt to respond as quickly as possible.

Before the silent-center syllables were presented, the experimenter explained that the syllables to be presented had been digitally processed such that the center of the syllables was silenced and not audible. This point was further illustrated by erasing the center of the *F*0 contours drawn on the blackboard. Similarly, before the onset-only syllables were presented, it was explained that the syllables had gone through further processing such that only the beginning of the syllables would be audible.

4. Data analysis

Response accuracy and reaction time were automatically recorded by BLISS. Reaction time was measured from stimulus offset to avoid the potential confound of intrinsic duration differences among the four tones. Only correct responses were included in the reaction time analysis.

Response data from all three types of syllables were first processed to evaluate the effects of musical background, syllable type, and tone. Specifically, analyses of variance (ANOVAs) were conducted on accuracy and reaction time with musical background (musicians and nonmusicians) as a between-subject factor, syllable type (intact, silent-center, and onset-only) and tone (1, 2, 3, and 4) as within-subject factors, and participants as a random factor. The data were then evaluated for each syllable type separately to examine the effects of tone and musical background. In particular, ANOVAs were conducted on accuracy and reaction time with tone as a within-subject factor, musical background as a between-subject factor, and participants as a random factor. When a main effect from the ANOVAs was significant, the Bonferroni *post hoc* test was used for pairwise means comparisons to keep the familywise type I error rate at 5%.

Missing cells in the reaction time data due to incorrect responses constituted 0.7% of the onset-only data for the musicians, 0.7% of the silent-center data for the nonmusicians, and 0.7% of the onset-only data for the nonmusicians. Separate ANOVAs were conducted with and without replacing the missing data with listener group averages. Since the two sets of ANOVAs showed the same patterns, only the ANOVAs with the replacements will be reported.

B. Results

1. All types of syllables considered

For response accuracy, ANOVAs revealed significant main effects of musical background $[F(1,70)=46.52,\ p<0.0001]$, syllable type $[F(2,140)=240.7,\ p<0.0001]$, and tone $[F(3,210)=26.94,\ p<0.0001]$. There were also significant interactions including syllable type × musical background $[F(2,140)=37.28,\ p<0.0001]$, tone × musical background $[F(3,210)=14.11,\ p<0.0001]$, syllable type × tone $[F(6,420)=26.09,\ p<0.0001]$, and syllable type × tone × musical background $[F(6,420)=4.22,\ p<0.0005]$.

Overall, accuracy was higher for the musicians (51%, SD=26) than for the nonmusicians (36%, SD=19). As expected, accuracy was highest for the intact syllables (56%, SD=24), followed by the silent-center syllables (45%, SD=24), and lowest for the onset-only syllables (29%, SD=14) [see the work of Gottfried (2007), who found no difference between intact and silent-center syllables]. All pairwise

means comparisons were significant. For the tone effect, accuracy was higher for tone 2 (49%, SD=22), tone 1 (47%, SD=26), and tone 4 (45%, SD=27) than for tone 3 (33%, SD=18). All pairwise means comparisons involving tone 3 were significant.

For reaction time, ANOVAs revealed significant main effects of syllable type [$F(2,140)=49.51$, $p<0.0001$] and tone [$F(3,210)=12.59$, $p<0.0001$]. There were also significant interactions including syllable type × musical background [$F(2,140)=4.22$, $p<0.05$], syllable type × tone [$F(6,420)=16.32$, $p<0.0001$], and syllable type × tone × musical background [$F(6,420)=2.24$, $p<0.05$].

In contrast to the accuracy results, no significant difference in reaction time was found between musicians (1240 ms, SD=473) and nonmusicians (1380 ms, SD=577). Consistent with the accuracy analysis, reaction time was shortest for intact syllables (1152 ms, SD=531), followed by silent-center syllables (1228 ms, SD=506), and longest for onset-only syllables (1549 ms, SD=471). All pairwise means comparisons were significant except for the contrast between intact and silent-center syllables. For the tone effect, reaction time was shortest for tone 1 (1231 ms, SD=468), followed by tone 3 (1249 ms, SD=526), tone 2 (1337 ms, SD=537), and tone 4 (1421 ms, SD=571). Significant pairwise means comparisons included tone 1-tone 2, tone 1-tone 4, and tone 3-tone 4.

Given the significant interactions, the main effects needed to be interpreted further. Inspection of the accuracy data arranged by syllable type and musical background revealed that the difference between the musicians and nonmusicians became less as the amount of acoustic input was reduced. For intact tones, the musicians outperformed the nonmusicians substantially. When acoustic input became minimal, musical background did not facilitate tone identification anymore [see the work of Gottfried (2007), who found a difference between musicians and nonmusicians only for silent-center syllables]. For reaction time, the pattern is consistent with the accuracy analysis, i.e., the contrast between the musicians and nonmusicians became less as the amount of acoustic input was reduced.

When arranged by tone and musical background, the accuracy data showed that the musicians outperformed the nonmusicians in the identification of all tones but tone 3, indicating that the musicians' advantage was neutralized for tone 3. The reaction time data, in contrast, showed that the musicians took less time than the nonmusicians in identifying all tones but tone 2. Overall, however, this pattern is consistent with the accuracy result.

When arranged by syllable type and tone, the accuracy data showed that tone 2 and tone 4 identifications were minimally compromised when syllable center was removed [see the work of Gottfried, 2007, who showed that tone 1 identification was compromised more by the missing center compared to the other tones]. Our interpretation is that tone 2 (low rising) and tone 4 (high falling) are contour tones with the beginning and ending $F0$'s being quite distinct from each other. In contrast, the beginning and ending $F0$ for tone 1 (high level) and tone 3 (low dipping) are at approximately the same height and less distinct from each other. Consequently, when only the onset and offset of a syllable were available, tones with more distinct $F0$ onset-offset pairs could be detected more easily. For reaction time, there was little contrast among the four tones for the silent-center and onset-only syllables. For the intact syllables, however, reaction time for tone 3 was particularly short compared to other tones. Since tone 3 could be associated with a creaky voice quality (Gottfried and Suiter, 1997; Liu and Samuel, 2004), listeners might be able to detect the qualitative difference earlier in the syllable than the $F0$ contour.

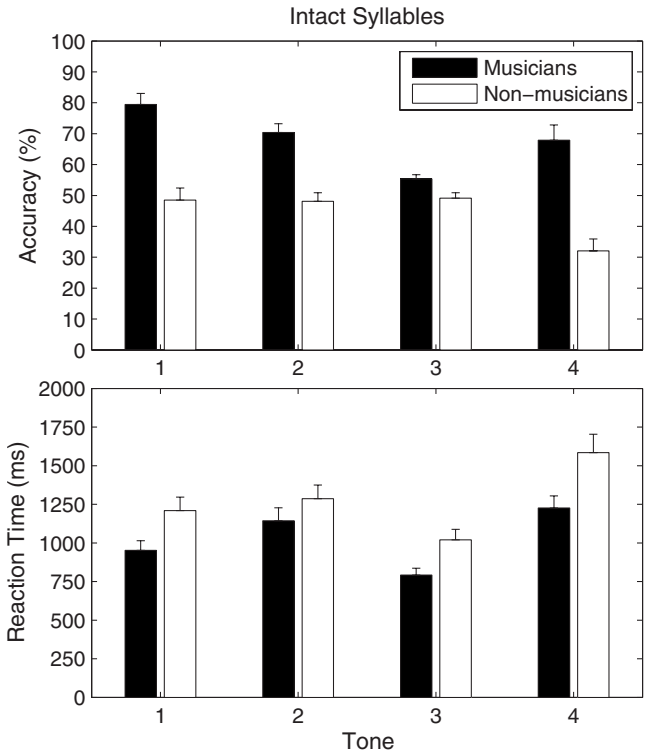


FIG. 2. Average accuracy and reaction time (+SE) of Mandarin tone identification as a function of musical background and tone for intact syllables ($n=36$).

To decipher the three-way interactions, we analyzed the accuracy and reaction time data for each type of syllables separately. The results are reported below.

2. Intact syllables

Figure 2 shows the accuracy and reaction time of intact tone identification as a function of musical background and tone. For accuracy, the ANOVA revealed significant main effects of musical background [$F(1,70)=51.36$, $p<0.0001$] and tone [$F(3,210)=11.35$, $p<0.0001$], and a significant musical background × tone interaction [$F(3,210)=11.56$, $p<0.0001$]. As predicted, the musicians (68%, SD=22) outperformed the nonmusicians (44%, SD=20). When both groups of listeners were considered, tone 1 (64%, SD=27) and tone 2 (59%, SD=20) were identified more accurately than tone 3 (52%, SD=10) and tone 4 (50%, SD=32). Significant contrasts included tone 1-tone 3, tone 1-tone 4, and tone 2-tone 4. The interaction shows that the accuracy difference between musicians and nonmusicians was substantial for tones 1, 2, and 4, but was small for tone 3.

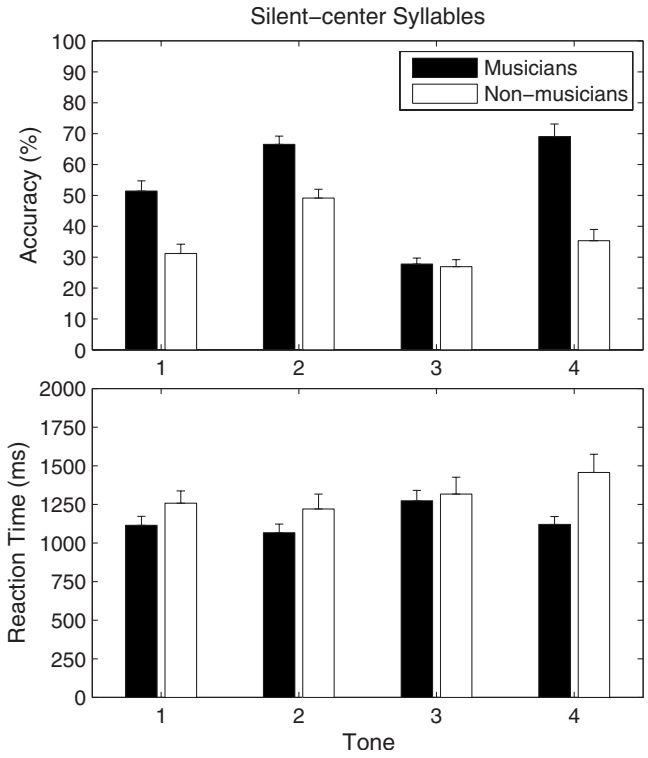


FIG. 3. Average accuracy and reaction time (+SE) of Mandarin tone identification as a function of musical background and tone for silent-center syllables ($n=36$).

For reaction time, the ANOVA revealed significant main effects of musical background [$F(1,70)=6.3$, $p<0.05$] and tone [$F(3,210)=37.34$, $p<0.05$]. There was no interaction. Consistent with the accuracy results, the musicians (1029 ms, SD=440) responded faster than the nonmusicians (1275 ms, SD=585). When both groups of listeners were considered, tone 3 (906 ms, SD=360) was identified most quickly, followed by tone 1 (1081 ms, SD=467), tone 2 (1215 ms, SD=516), and tone 4 (1405 ms, SD=625). All pairwise comparisons were significant.

In summary, tone identification performance for the intact syllables was quite impressive considering the brief instructions on the novel linguistic contrast in the unfamiliar language, the speaker variability in the stimuli, and the pressure to respond quickly. Although both the musicians and nonmusicians identified the tones with accuracy exceeding chance (25%), the musicians were more accurate and faster than the nonmusicians. This result is consistent with the idea that musical training facilitates linguistic tone perception.

3. Silent-center syllables

Figure 3 shows the accuracy and reaction time of tone identification for the silent-center syllables. For accuracy, the ANOVA revealed significant main effects of musical background [$F(1,70)=47.87$, $p<0.0001$] and tone ($F(3,210)=46.66$, $p<0.0001$), and a significant musical background $\times$tone interaction [$F(3,210)=11.75$, $p<0.0001$]. With the middle 70% of the voiced portion of a syllable silenced, the musicians (54%, SD=25) still outperformed the nonmusicians (36%, SD=20). When both groups of listeners were considered, tone 2 (58%, SD=19) and tone 4 (52%, SD=29) were identified more accurately than tone 1 (41%, SD=21) and tone 3 (27%, SD=13). All pairwise means comparisons were significant except for the contrast between tone 2 and tone 4. The interaction shows that the accuracy difference between musicians and nonmusicians remained substantial for tones 1, 2, and 4, but was minimal for tone 3. This is the same pattern as what was found for the intact syllables.

For reaction time, the ANOVA revealed a significant main effect of tone [$F(3,210)=3.61$, $p<0.05$]. There was no effect of musical background or interaction. When both groups of listeners were considered, tone 2 (1143 ms, SD=475) was identified most quickly, followed by tone 1 (1186 ms, SD=422), tone 4 (1288 ms, SD=572), and tone 3 (1295 ms, SD=536). Only the tone 2-tone 3 contrast was significant. Although musicians (1144 ms, SD=354) on average responded faster than nonmusicians (1312 ms, SD=612), the difference was not significant.

In summary, identification accuracy dropped and reaction time increased as acoustic input was reduced. The musicians were still more accurate, but no longer faster than the nonmusicians. The average identification accuracy (54%) for the musicians remained above chance despite missing the syllable center. Identification accuracy for tone 2 and tone 4 was hardly compromised compared to intact tone 2 and tone 4. As noted, our interpretation was that tone 2 (low rising) and tone 4 (high falling) have more distinct beginning $F0$ and ending $F0$ compared to tone 1 (high level) and tone 3 (low dipping). When the onset and offset of a syllable were available, tones with more distinct onset-offset $F0$ pairs could be detected more easily, contributing to the higher identification accuracy. This finding also indicated potential benefits of highlighting the beginning and end of these two tones for listeners who have not had extensive experience with the language.

4. Onset-only syllables

Figure 4 shows the accuracy and reaction time of tone identification for the onset-only syllables. The ANOVA revealed a significant main effect of tone [$F(3,210)=17.04$, $p<0.0001$] and a significant musical background $\times$tone interaction [$F(3,210)=3.79$, $p<0.05$]. Although the musicians (31%, SD=15) on average still outperformed the nonmusicians (28%, SD=14), the difference was not significant [$F(1,70)=3.57$, $p=0.06$]. When both groups of listeners were considered, tone 1 (35%, SD=18), tone 4 (32%, SD=13), and Tone 2 (31%, SD=12) were identified more accurately than tone 3 (20%, SD=11). All pairwise means comparisons involving tone 3 were significant. The interaction shows that the musicians were more accurate than the nonmusicians only for tones 1 and 4.

For reaction time, the ANOVA revealed a significant main effect of tone [$F(3,210)=6.91$, $p<0.0005$] and a significant musical background$\times$tone interaction [$F(3,210)=2.73$, $p<0.05$]. As with the accuracy analysis, there was no difference between the musicians and nonmusicians. When both groups of listeners were considered, tone 1 (1427 ms, SD=451) was identified most quickly, followed by tone 3 (1544 ms, SD=460), tone 4 (1571 ms, SD=480), and tone

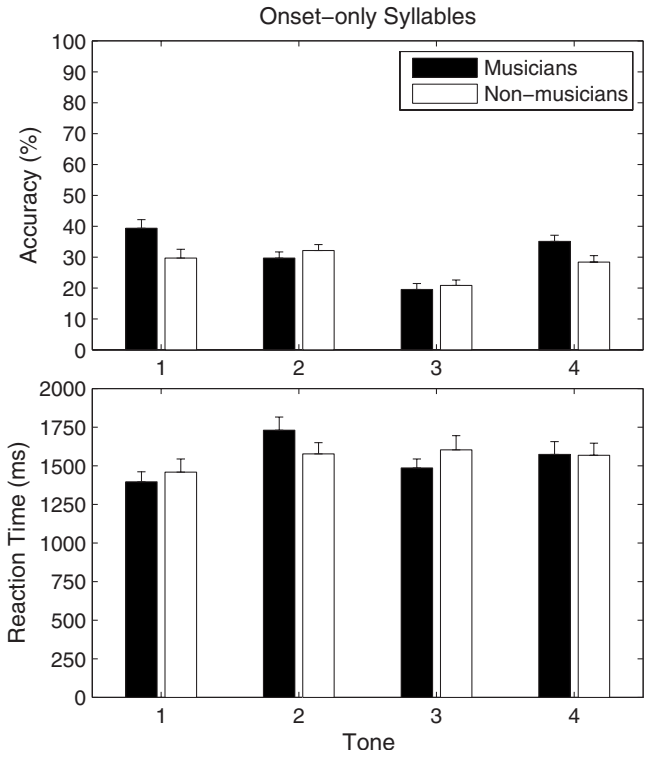


FIG. 4. Average accuracy and reaction time (+SE) of Mandarin tone identification as a function of musical background and tone for onset-only syllables ($n=36$).

2 (1653 ms, SD=477). Significant contrasts included tone 1-tone 4 and tone 1-tone 2. Consistent with the accuracy analysis, the interaction shows minimal difference between the musicians and the nonmusicians for Tones 1 and 4.

In summary, identification accuracy dropped further as acoustic input was reduced to onset only. In contrast to the previous two syllable types, both accuracy and reaction time for the onset-only syllables were comparable between the musicians and nonmusicians. For both groups of listeners, identification accuracy fell close to chance level. Unlike the silent-center syllables, where listeners might be able to reconstruct tone shapes based on the onset and offset $F0$, the interpolation strategy would not be useful when only the onset was present. $F0$ contour information is also unlikely to be useful for tone judgments given the brevity of the onset-only syllables (Greenberg and Zee, 1979). Although $F0$ height could serve to broadly classify the four tones into high-onset (tones 1 and 4) and low-onset (tones 2 and 3) tones, the high-low judgment is necessarily a relative one and would require estimation of a speaker's $F0$ range as a reference frame. However, the presence of multiple speakers in the stimulus set might have added to the challenge of reliable $F0$ range estimation, which could have further contributed to the overall low accuracy.

C. Acoustic analyses

The Mandarin tone experiment revealed several noteworthy response patterns regarding the effects of musical background, syllable type, and specific tones. To interpret these patterns, we have referred to the nature of the acoustic input, e.g., the type of $F0$ information in the acoustic signal and the differences among the tones. To validate these interpretations, acoustic analyses were conducted on the duration and $F0$ of the stimuli to verify that the acoustic contrasts were actually present. These two measures were chosen because duration is a direct measure of the amount of acoustic input, and $F0$ has been established to be the primary acoustic correlate of Mandarin tones.

TABLE I. Average duration (in ms) of the onset/offset and center of the voiced portion of the syllables. The onset constituted 15%, center 70%, and offset 15% of the voiced portion of the syllables. Standard deviations are shown in parentheses.

Speaker	Tone	Onset/Offset	Center
Female	1	58 (12)	279 (58)
	2	59 (15)	274 (68)
	3	65 (23)	304 (110)
	4	34 (8)	159 (37)
Male	1	42 (7)	197 (35)
	2	42 (9)	198 (43)
	3	45 (17)	210 (77)
	4	27 (5)	127 (25)

1. Duration

Table I shows the average duration of the three components (onset, center, and offset) of the voiced portion of the syllables. Recall that the onset and offset each constituted 15% of the voiced portion and the center included 70% of the voiced portion. An ANOVA was conducted on the duration of the onset with speaker gender (female and male) and tone (1, 2, 3, and 4) as fixed factors and speakers as a random factor. No further ANOVAs were conducted on center or offset duration separately since they were proportional to the onset duration and would have generated the same results.

For duration, the ANOVA revealed significant main effects of speaker gender [$F(1,120)=39.35$, $p<0.0001$] and tone [$F(3,120)=21.32$, $p<0.0001$]. There was no interaction. Onset duration was longer for the female speakers (54 ms, SD=19) than for the male speakers (39 ms, SD=12). When both female and male speakers were considered, the onset duration for tone 3 (55 ms, SD=22), tone 2 (51 ms, SD=15), and tone 1 (50 ms, SD=13) was longer than tone 4 (31 ms, SD=8). All pairwise contrasts involving tone 4 was significant. Center duration was likewise longer for the female speakers (252 ms, SD=91) than for the male speakers (183 ms, SD=58). When both female and male speakers were considered, the center duration for tone 3 (257 ms, SD=105),tone 2 (236 ms, SD=68), and tone 1 (233 ms, SD=60) was longer than tone 4 (143 ms, SD=35).

In summary, the tones produced by the female speakers were on average longer than those produced by the male speakers, indicating that the female speakers were speaking at a slower rate. In addition, tone 4 was consistently the shortest tone, which is consistent with the literature on the intrinsic duration difference among the four tones (e.g., Tseng, 1981). Importantly, duration did not predict the relative identification accuracy in the task. In particular, tone 4 was the shortest and thus had the least amount of acoustic input, but it was often identified most accurately. On the

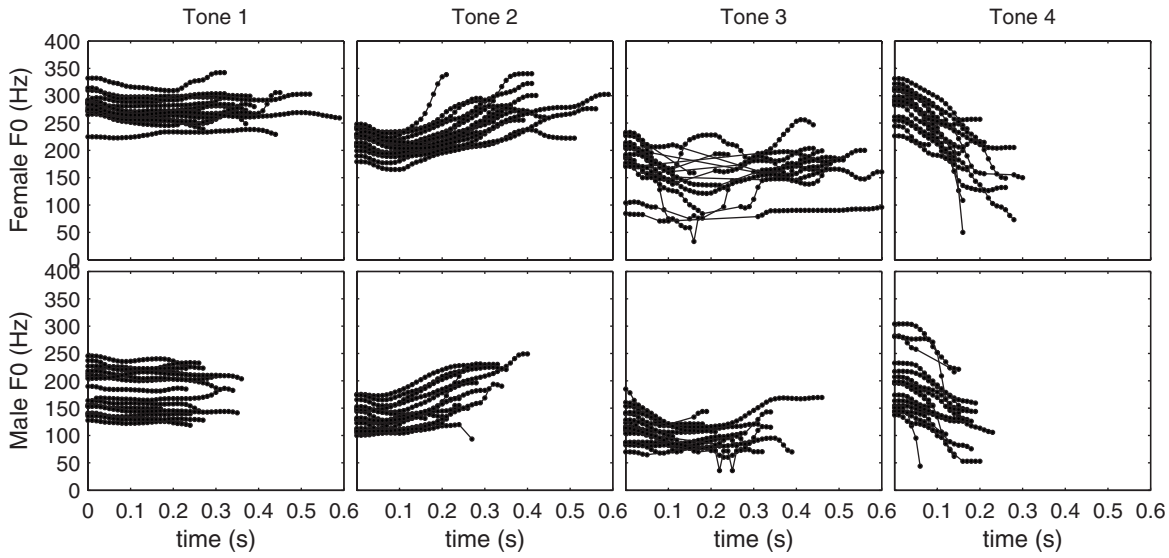


FIG. 5. The *F*0 contours of the four Mandarin tones produced by the 16 female (top) and 16 male (bottom) speakers.

other hand, even though tone 3 had the longest duration among the four tones, it was consistently identified with the lowest accuracy. Therefore, duration alone was not a good predictor of Mandarin tone identification performance. It should be noted, however, that intrinsic duration difference has been shown to be a useful cue to Mandarin tone identification for native listeners (Blicher *et al.*, 1990; Liu and Samuel, 2004; Whalen and Xu, 1992; Xu *et al.*, 2002).

2. Fundamental frequency

Figure 5 shows the *F*0 contours of the four tones produced by all speakers, shown separately for the female and male speakers. Despite obvious variability among the speakers, the *F*0 contours are generally consistent with traditional descriptions, i.e., high level for tone 1, low rising for tone 2, low dipping for tone 3, and high falling for tone 4. Recall from the perception experiment that identification of silent-center tone 2 and tone 4 was hardly compromised compared to intact tone 2 and tone 4. The perceptual resistance to the silent-center modification was attributed to their distinct onset-offset *F*0 values. Our acoustic data here are consistent with the interpretation. Finally, the difference in *F*0 range between female and male speakers was expected. That is, the female speakers had a higher *F*0 range than the male speakers.

D. Summary and discussion

With brief instructions and under time pressure, musicians without prior Mandarin experience were able to identify Mandarin tones with 68% accuracy, exceeding chance (25%). When the majority of the voiced portion of the syllables was digitally silenced and devoid of *F*0 information, the musicians managed to identify the tones with 54% accuracy, still exceeding chance. The accuracy dropped to 31% when the acoustic input was further reduced to the beginning 15% of the voiced portion of the syllable, indicating syllable onset alone was not sufficient for reliable tone identification.

With identical instructions, age- and gender-matched nonmusicians did not perform as well as the musicians. In particular, the nonmusicians identified the intact, silent-center, and onset-only syllables with accuracies at 44%, 35%, and 27%. The overall higher identification accuracy by the musicians indicates that listeners with a musical background had an advantage in lexical tone identification. This observation is consistent with studies showing superior performance in Mandarin tone perception by English listeners with musical training (Gottfried, 2007; Gottfried and Riester, 2000; Gottfried *et al.*, 2001; Alexander *et al.*, 2005; Wong *et al.*, 2007).

Nonetheless, the contrast in identification accuracy alone did not address the nature of the processing difference between the musicians and nonmusicians. In the current study, comparisons among the three types of syllables and among the four tones further revealed similarities and differences between the musicians and nonmusicians in perceiving Mandarin tones. Both groups showed a linear decrease in accuracy and increase in reaction time as acoustic input was reduced. Both groups also showed little drop in accuracy for silent-center tone 2 and tone 4 compared to intact tone 2 and tone 4, indicating that the ability to reconstruct tone identity based on onset-offset information was present in both groups of listeners.

On the other hand, the musicians were different from the nonmusicians in how the tone identification performance was influenced by the acoustic modification. Specifically, the advantage of musicians became less obvious when acoustic input was reduced. When tonal information was fully present in the intact syllables, the musicians were both more accurate and faster in tone identification [see the work of Gottfried, (2007), who found no accuracy difference between musicians and nonmusicians]. When the majority of the *F*0 information was unavailable in the silent-center syllables, the musicians were still more accurate but were no longer faster [see the work of Gottfried, 2007, who also showed an accuracy difference between musicians and nonmusicians]. Finally, when *F*0 information was limited to the beginning

15% of the voiced portion of the syllables, the musicians were neither more accurate nor faster than the nonmusicians. In other words, this experiment revealed that the musicians' processing advantage in tone identification depends on the type of input available in the acoustic signal. Music background did help in lexical tone identification, but the advantage was compromised by the reduction of acoustic input.

While these results revealed further information about lexical tone perception by non-native musicians and nonmusicians, the question remained of what specific component of musical background contributed to the Mandarin tone identification performance. As noted, both *F*0 direction and *F*0 height are involved in Mandarin tone distinctions. It could be the enhanced sensitivity to *F*0 direction/movement (Gottfried, 2007), the ability to detect *F*0 height, or both, that contributed to the musicians' better identification. Although the musicians as a group outperformed the nonmusicians, there was considerable variability among the musicians in their lexical tone identification performance. On the other hand, it is possible that some variability could also exist in the musicians' ability to process musical pitch information. Therefore, it would be of interest to explore whether the performance in lexical tone identification is correlated with performance in musical pitch perception.

As noted, Gottfried and Riester (2000), also reported in Gottfried, 2007, showed higher accuracy by musicians in identifying both Mandarin tones and sine-wave tone glides with analogous *F*0 shifts, suggesting that the ability to perceive *F*0 direction underlies the superior performance by musicians in Mandarin tone identification. In the next experiment, we evaluate the potential role of *F*0 height detection in Mandarin tone identification. To that end, we adopted a task that gauges the musicians' ability to name a musical note without a reference pitch, i.e., absolute pitch (Deutsch *et al.*, 2006), as an index of musical pitch processing. As noted, absolute pitch has been implicated in the acquisition and processing of lexical tones (Deutsch *et al.*, 2004a; Deutsch *et al.*, 2006). If absolute pitch is indeed involved in lexical tone perception, we would expect that musicians with higher scores in the absolute pitch task would also perform better in the lexical tone identification task.

III. EXPERIMENT 2: MUSICAL NOTE IDENTIFICATION BY MUSICIANS

In this experiment, the 36 musicians who participated in the Mandarin tone experiment were asked to listen to synthesized musical tones of three timbres (pure tone, piano, and viola) and to identify the notes in the absence of a reference pitch. The setup of the experiment was identical to the absolute pitch task used in Deutsch *et al.*, 2006 except for the use of two additional timbres in the stimuli. The use of multiple timbres was motivated by the finding that the manifestation of absolute pitch could depend on the specific instruments used to generate the musical notes (Lockhead and Byrd, 1981). It is therefore informative to evaluate the potential impact of timbre on musical note identification.

A. Method

1. Materials

Thirty-six notes that spanned a three-octave range from C_3 (131 Hz) to B_5 (988 Hz) with the equal-tempered scale were synthesized with three timbres (pure tone, piano, and viola) for a total of 108 notes. The pure tones were synthesized using the MATLAB software (The MathWorks). The piano and viola notes were synthesized using a Kurzweil K2000 synthesizer tuned to the standard A_4 at 440 Hz. The duration for all notes was 500 ms.

For each timbre, the 36 notes were ordered such that any two consecutive notes were separated by more than an octave. As noted by Deutsch *et al.* (2006), this was done to prevent listeners from developing relative pitch as a reference for the task. The 36 notes of a given timbre were divided into three blocks of 12 notes, with a 5 s interstimulus interval and a 10 s break between the blocks. Notes of a given timbre were always presented together. To counterbalance the order of timbre presentation, six lists of stimulus presentation were created: (1) pure tone-piano-viola; (2) pure tone-viola-piano; (3) piano-pure tone-viola; (4) piano-viola-pure tone; (5) viola-pure tone-piano, and (6) viola-piano-pure tone. The 36 musician participants were randomly assigned to receiving one of the six lists such that a given list was used for six participants.

2. Participants

The 36 musicians in the Mandarin tone identification experiment participated in this experiment.

3. Procedure

The synthesized musical notes were saved as individual audio files and imported to AVRunner for stimulus presentation. Twelve practice trials, synthesized as oboe sounds with the Kurzweil K2000 synthesizer, were given prior to the actual experiment to familiarize the participants with the presentation format. The practice notes were selected from the same pitch range (C_3-B_5) and were also 500 ms long. As in the actual experiment, any two consecutive notes were separated by more than an octave. None of the practice notes appeared in the actual experiment. No feedback was given.

Participants listened to the stimuli through a pair of headphones connected to a Windows personal computer. They were instructed to write down the notes on a customized staff paper. In particular, the participants were told that they would be listening to nine blocks of 12 notes of three timbres ranging from C_3 to B_5. Their task was to notate the notes that they had heard on the staff paper immediately after each note was played and to apply accidental signatures if applicable. The participants were also told that there would be 5 s to respond to each note and that there would be a 10 s break between blocks.

4. Data analysis

The written responses were graded by the second author of this study (a musician). One-way repeated measures ANOVAs were conducted on response accuracy with timbre (pure tone, piano, and viola) as a within-subject factor and

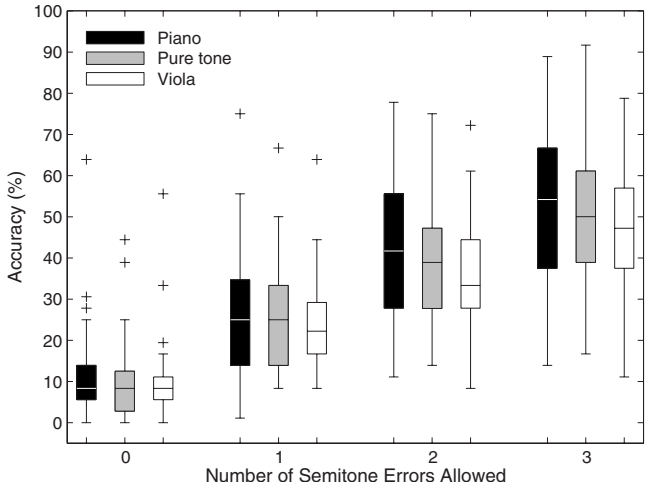


FIG. 6. A box plot showing the accuracy of musical note identification for piano, pure tone, and viola as a function of the number of semitone errors allowed.

participants as a random factor. There were three dependent variables, including (1) percentage of accurate note identification, (2) percentage of accurate note identification allowing one-semitone errors, (3) percentage of accurate note identification allowing two-semitone errors, and (4) percentage of accurate note identification allowing three-semitone errors. When a main effect from the ANOVAs was significant, the Bonferroni *post hoc* test was used for pairwise means comparisons to keep the familywise type I error rate at 5%.

The motivation for allowing up to three-semitone errors in the dependent measures was to avoid potential floor-level performance across the board due to the challenging nature of the task. In particular, Deutsch *et al.*, (2006) found that only 15% of the English-speaking musicians at the Eastman School of Music were able to attain identification accuracy of 85% or higher in a similar task. If correct identification rate was uniformly low and individual variability was completely absent, the correlation analyses to be performed between music note and lexical tone identification would not be fully meaningful. Our purpose was not to evaluate the percentage of musicians with absolute pitch per se, but rather to obtain a measure of the musicians' absolute pitch ability for the correlation analyses.

B. Results

Figure 6 shows a box plot for the accuracy of musical note identification arranged by timbre and the number of semitone errors allowed. When an exact match was required, accuracy was uniformly low for all three timbres: pure tone (mean=12%, SD=12), piano (mean=11%, SD=10), and viola (mean=11%, SD=10). Since there are 12 semitones within an octave, chance level performance is 8.3%. The obtained averages therefore indicate that the musicians' identification accuracy was within chance level. Inspection of individual data showed that the highest accuracy attained was 64%, i.e., none of the musicians met the 85% criterion for absolute pitch as defined in Deutsch *et al.* (2006). The ANOVA revealed no significant timbre effect.

When one-semitone errors were allowed, accuracy was expectedly higher: piano (mean=26%, SD=17), pure tone (mean=26%, SD=14), and viola (mean=24%, SD=11). Although these averages appear to be improved, they are still within chance level. In particular, permitting one-semitone errors effectively allows a range of three semitones, indicating a chance level of 25%. Inspection of individual data showed that the highest accuracy attained was 75%, i.e., none of the musicians met the 85% criterion for absolute pitch. The ANOVA revealed no timbre effect.

When two-semitone errors were allowed, accuracy became higher as expected: piano (mean=42%, SD=18), pure tone (mean=40%, SD=16), and viola (mean=35%, SD=14). Nonetheless, these averages are still within chance level. Specifically, permitting two-semitone errors effectively allows a range of five semitones, indicating a chance level of 41.7%. Inspection of individual data showed that the highest accuracy attained was 78%. The ANOVA revealed a significant timbre effect $[F(2,70)=3.41,\ p<0.05]$. Pairwise means comparisons showed the accuracy for piano was higher than that for viola.

When three-semitone errors were allowed, accuracy became expectedly higher: piano (mean=53%, SD=19), pure tone (mean=51%, SD=18), and viola (mean=47%, SD=17). As before, although these averages appear to be further improved, they are still within chance level. In particular, permitting three-semitone errors effectively allows a range of seven semitones, indicating a chance level of 58.3%. Inspection of individual data showed that the highest accuracy attained was 92%. Four of the 36 musicians identified the notes at 85% correct or higher. The ANOVA revealed no significant timbre effect.

C. Summary and discussion

The task of identifying musical notes without a reference pitch was apparently challenging. By the criterion used in Deutsch *et al.* (2006), none of the musicians participating in this experiment qualified as possessing absolute pitch. Even when the scoring criterion was relaxed to allow errors as many as three semitones, average identification performance remained at chance level. Recall that the average age at which our musicians first received their music training was 9.4 (SD=3.2). Inspection of the data in Deutsch *et al.*, 2006 (Fig. 1) revealed that none of their English-speaking musicians in the corresponding age (of commencement of musical training) group possessed absolute pitch either. Our result is therefore consistent with the idea that a critical period is involved in the acquisition of absolute pitch.

Overall, timbre did not make a difference except when two-semitone errors were allowed. In this case, accuracy was higher for piano sounds than for viola sounds. This result is perhaps not surprising considering that all our musicians were required by their program of study to demonstrate proficiency in piano irrespective of their major instrument. Consequently, on average, they are likely to have more experience with piano than with viola.

Finally, individual variability can be clearly seen from the box plot, more prominently when the scoring criterion

was relaxed to allow more errors. However, on average, performance in the musical note identification task was still within chance level, indicating none of the musicians used in the current study could be claimed to possess absolute pitch. The absence of absolute pitch in the musicians posed a challenge to the interpretation of the planned correlation analyses between the accuracy measures in the Mandarin tone identification task (response accuracy for intact, silent-center, and onset-only syllables) and those in the musical note identification task (accuracy of note identification allowing zero-, one-, two-, and three-semitone errors). Even though our preliminary analyses showed only 1 significant correlation out of the 12 possible correlations ($r=0.329$, $p<0.05$, found between intact syllables identification and musical note identification when three-semitone errors were allowed), a different pattern could emerge for musicians who do possess absolute pitch.

IV. GENERAL DISCUSSION

Two perception experiments investigated the processing of linguistic and musical pitch by musicians and nonmusicians. The Mandarin tone experiment showed that the musicians were able to identify multispeaker intact and silent-center Mandarin tones with accuracy exceeding chance. The musicians also outperformed age- and gender-matched nonmusicians, even though the group difference became smaller as acoustic input was reduced. In the musical pitch task given to the musicians, accuracy of musical note identification remained within chance level even when errors up to three semitones were allowed, indicating none of the musicians used in this study possessed absolute pitch as defined in Deutsch *et al.*, 2006.

Although the finding that musicians outperformed nonmusicians in lexical tone perception is not new (Alexander *et al.*, 2005; Gottfried, 2007; Gottfried and Riester, 2000; Gottfried *et al.*, 2001; Wong *et al.*, 2007), the present study revealed new information about the nature of lexical tone perception by non-native listeners with or without a musical background. In particular, by simulating common challenges to speech perception with degraded stimuli and speaker variability, the present study showed that the benefit of musical experience varied depending on the type of acoustic input. Specifically, even though the musicians were faster and more accurate in identifying intact Mandarin tones, the reaction time advantage disappeared for silent-center syllables. When acoustic input was further reduced to onset-only syllables, the accuracy advantage also disappeared. In other words, the advantage of musical training in learning to identify nonnative lexical tones depends on the type of acoustic input. It would be of interest to investigate if other types of degraded signals (e.g., signals presented with noise) would exert a similar effect.

In addition, the group-tone interaction found in all three types of syllables (Figs. 2–4) showed that the musicians' advantage arose primarily from tones 1, 2, and 4. In contrast, tone 3 identification accuracy was comparable between the two groups of listeners. The musicians were particularly error prone in tone 3 identification. For the nonmusicians, tone 3 also generated the lowest accuracy among the four tones in silent-center and onset-only syllables. Taken together, it appears that tone 3 was generally difficult to identify for both groups. The difficulty could be attributed to the variable nature of tone 3. Although phonologically specified as a low-dipping tone with a falling-rising contour, the rising part does not normally show up phonetically except in isolation or in a phrase-final position. In addition, tone 3 is also involved in a common tone sandhi where tone 3 becomes tone 2 (a rising tone) when followed by another tone 3. Finally, for many speakers, there is often glottalization in the middle of the tone, resulting in breaks in the *F*0 contour (Gottfried and Suiter, 1997; Liu and Samuel, 2004). Although the tone 3 tokens used in this study were all produced in isolation, the acoustic data (Fig. 5) showed that not all tone 3 tokens maintained its canonical *F*0 contour and that there were visible *F*0 breaks for some speakers. The variable nature of this tone could have contributed to the difficulty in identification and the reduced advantage of the musicians.

Despite these contrasts, there were some similarities between the musicians and nonmusicians. It is perhaps not surprising that both groups showed a linear drop in accuracy and a corresponding increase in reaction time as a function of the type of acoustic input. More interestingly, analysis by tone revealed that the four tones were influenced differently by the acoustic reduction. In particular, for both groups of listeners, tone 2 and tone 4 identifications were minimally affected by the removal of the syllable center. That is, both groups were able to uncover the identity of these two tones despite the missing center. It was hypothesized that the distinct onset and offset *F*0 for tone 2 and tone 4 would be particularly useful for reconstructing pitch contour when the syllable center was missing. This observation was supported by our acoustic data (Fig. 5).

Nonetheless, the overall accuracy for the musicians (68% for intact tones and 54% for silent-center tones) is quite remarkable considering the minimal instructions given, the novel linguistic contrast in the unfamiliar language, the speaker variability, and the time pressure to respond quickly. In a study with a similar speeded-response task by Lee *et al.* (2006), 40 English-speaking listeners with one to three years of Mandarin experience achieved accuracy of 76% for intact tones and 59% for silent-center tones (see native listeners in Lee *et al.*, 2008a achieved 97% for intact tones and 86% for silent-center tones). Gottfried and Suiter (1997) in a study with a similar task without speeded response reported average accuracy of 79% for intact tones and 65% for silent-center tones by nine English-speaking listeners with an average of five years of Mandarin experience (see native listeners in the same study achieved 98% for intact tones and 90% for silent-center tones). Importantly, the stimuli in Lee *et al.*, 2006 or Gottfried and Suiter, 1997 were produced by a single speaker, i.e., the participants in those two studies did not have to deal with speaker variability. With substantially less Mandarin experience but the added challenges, the musicians' tone identification performance in the present study is quite impressive.

What could be the basis for the superior performance by the musicians in tone identification? We attempted to evalu-

ate the hypothesis that absolute pitch is implicated in lexical tone processing (Deutsch *et al.*, 2004a; Deutsch *et al.*, 2006) by examining the correlation between Mandarin tone identification and absolute pitch performance. However, the absence of absolute pitch in our musician participants made it impossible to fully evaluate the correlation results. In particular, the lack of robust correlation in the current data cannot rule out the potential association between absolute pitch and lexical tone perception. Significant correlations between Mandarin tone identification and musical note identification may be obtained if the Mandarin tone task is given to musicians who actually possess absolute pitch.

In addition, the nature of the stimuli and response format may have encouraged the participants to attend more to *F*0 contour information than *F*0 height information. In particular, since a large number of speakers were used to generate the stimuli, there was considerable overlap in *F*0 range for different tones (Fig. 5). Given the variability of *F*0 height introduced by the use of multiple speakers, listeners may be inclined to use contour rather than height information. In addition, even though the response keys were arranged to reflect the relative height of the tones and the listeners were instructed to pay attention to both contour and height information, the keys did show tonal contours prominently, which could have primed the listeners to focus more on contour information.

Furthermore, the benefit of absolute pitch may be revealed in a task that evaluates the acquisition of lexical tones after substantial training. In particular, learning a novel linguistic contrast for adult learners usually needs extensive time and exposure to the language. The present study, however, did not allow ample opportunities for the participants to get familiar with the Mandarin tonal contrasts. What was evaluated in the present study was the rapid perceptual learning rather than the acquisition of lexical tones. It is possible that individuals with absolute pitch could demonstrate an advantage in the acquisition of lexical tones with more training.

Even though the musicians used in this study did not possess absolute pitch, they did achieve accuracy values that substantially exceeded chance for intact and silent-center Mandarin tones, indicating a certain degree of success in learning to map the acoustic input onto Mandarin tonal categories. Their performance also approached that of the nonnative learners who had more Mandarin experience and who did not have to deal with speaker variability (Gottfried and Suiter, 1997; Lee *et al.*, 2006). Furthermore, the musicians did outperform the nonmusicians in all but the onset-only syllable task. All these findings indicate that musical experience indeed facilitated lexical tone perception. Since none of the musicians in this study possessed absolute pitch, the advantage must have come from other sources. Gottfried and Riester (2000), also reported in Gottfried, 2007, to our knowledge the only study that used both linguistic and nonlinguistic pitch tasks, suggested that musicians' superior performance in lexical tone identification may be attributed to the ability to track pitch movement effectively. This observation is consistent with recent neurophysiological evidence showing more faithful tracking of pitch by musicians than by nonmusicians (Wong *et al.*, 2007). In other words, the musicians' advantage in the present study may have come from their superior ability to track and remember the tonal patterns and to identify new examples of the patterns. More broadly, the literature on music and language has also identified association between musical pitch discrimination and reading abilities in 5-year-olds (Anvari *et al.*, 2002) and association between pitch pattern perception and second-language skills (Slevc and Miyake, 2006). It would be of interest for future studies to explore the association between lexical tone processing and musical processing abilities other than absolute pitch such relative pitch detection and sensitivity to duration differences.

Although the present study did not provide evidence for the association between lexical tone perception and absolute pitch, the absolute pitch—lexical tone link appears well motivated by other evidence. As Deutsch *et al.* (2004b) noted, both lexical tone and absolute pitch involve associating a particular or several pitches with a verbal label. Deutsch and co-workers also showed that a speaker's language background, linguistic experience in childhood, and vocal range are associated with the tritone paradox (Deutsch, 1991; Deutsch *et al.*, 2004b; Deutsch *et al.*, 1990). Dolson (1994) reported that the pitch range of an individual's speaking voice is constrained primarily by his or her linguistic community, suggesting that listeners could have formed certain pitch templates through exposure to the prevailing pitch range of their linguistic community. Consequently, they are able to use those templates in the processing of vocal pitch. This idea seems consistent with Honorof and Whalen (2005), who reported that English listeners can locate *F*0 height reliably from isolated and naturally produced speech without context or prior exposure to a speaker's *F*0 range. Finally, there is evidence showing the use of absolute pitch by infants in perceptual learning (Saffran and Griepentrog, 2001). The acquisition of absolute pitch is also associated with early musical training, implying a critical period for absolute pitch acquisition (Deutsch *et al.*, 2006). These parallels led to the speculation that through exposure to the lexically contrastive use of pitch, an implicit form of absolute pitch is acquired by tone language speakers (Deutsch, 2006).

However, although absolute pitch and lexical tone both involve associating pitch with a verbal label, there are important differences between these two constructs. Absolute pitch requires associating a particular pitch with a verbal label without a reference pitch. Lexical tone, on the other hand, normally requires associating a verbal label with a pitch *pattern*, not just a single pitch. This is true at least for languages with contour tones including Mandarin, where pitch contour plays a primary role in tone identity. It would be of interest for future studies to investigate the role of absolute pitch in processing tones in languages with level tones, which presumably would rely critically on the detection of pitch level.

Furthermore, although *F*0 is the primary acoustic correlate of lexical tones, it is not the only acoustic cue that listeners could use to uncover tone identity. As with all phonetic contrasts, there are multiple acoustic cues to lexical tone contrasts including intrinsic duration and amplitude

(e.g., Whalen and Xu, 1992), which explains why whispered tones could be identified despite missing *F*0 information completely (Liu and Samuel, 2004). In other words, to identify lexical tones is to process all relevant information from the output of a vocal tract and map that information onto linguistically significant categories. Although pitch is clearly implicated in lexical tone processing, processing lexical tone as a linguistic contrast is not constrained to processing pitch information.

Finally, just as they do for consonants and vowels, tone language speakers use lexical tones to encode a linguistic message, and listeners decode acoustic patterns to uncover the tones in order to decipher a linguistic message. That is, as with any phonetic distinctions, lexical tones are linguistic contrasts. The association between an acoustic pattern and a verbal label can be acquired without explicit instruction by normally developing native speakers of a tone language. In contrast, the verbal label associated with absolute pitch (i.e., note names) has to be taught. While pitch processing is no doubt implicated in both tasks, the functional contrast in whether pitch is interpreted linguistically could be a critical difference (e.g., Remez *et al.*, 1981). The extent to which common processing mechanisms are involved in both domains awaits future research.

V. CONCLUSION

The present study contributes to the literature on language-music relationship in several ways. The Mandarin tone identification experiment showed that the advantage of musical training in lexical tone perception depends on the nature of the acoustic input. When the amount of tonal information was reduced, the identification performance difference between the musicians and nonmusicians also became less. Furthermore, the acoustic modifications also revealed tone-specific effects of reduction; namely, tone 2 and tone 4 were minimally compromised even when the majority of the syllable center was absent. Finally, by administering both the Mandarin tone and the absolute pitch task on the same musicians, we attempted to evaluate the association between these two tasks and thus the hypothesis that absolute pitch is implicated in lexical tone perception. Although our data could not fully evaluate the correlations between the two tasks due to the absence of absolute pitch in the musicians, the musicians clearly outperformed the nonmusicians in Mandarin tone identification, suggesting that abilities other than absolute pitch contributed to the musicians' superior performance in Mandarin tone identification. Future work could further explore the basis of the advantage.

ACKNOWLEDGMENT

We are grateful to Diana Deutsch and an anonymous reviewer for their valuable comments. We also thank the School of Music at Ohio University for making the Kurzweil K2000 synthesizer available for stimulus construction and Fuh-Cherng Jeng for synthesizing the pure tone stimuli. We thank Sarah Letsky for assistance in administering the experiments, Ning Zhou, En Ye, Anne Marie Christy, and Gayatri Ram for assistance in data processing, and Li Xu and Z. S. Bond for discussions. This research was partially supported by professional development funds from the School of Hearing, Speech and Language Sciences and the Honors Tutorial College at Ohio University.

Alexander, J., Wong, P. C. M., and Bradlow, A. (**2005**). "Lexical tone perception in musicians and nonmusicians," Proceedings of Interspeech' 2005–Eurospeech–Ninth European Conference on Speech Communication and Technology, Lisbon, Portugal.

Anvari, S. H., Trainor, L. J., Woodside, J., and Levy, B. A. (**2002**). "Relations among musical skills, phonological processing, and early reading ability in preschool children," J. Exp. Child Psychol. **83**, 111–130.

Blicher, D. L., Diehl, R. L., and Cohen, L. B. (**1990**). "Effects of syllable duration on the perception of the Mandarin tone 2/tone 3 distinction: Evidence of auditory enhancement," J. Phonetics **18**, 37–49.

Burnham, D., Peretz, I., Stevens, K., Jones, C., Schwanhäusser, B., Tsukada, K., and Bollwerk, S. (**2004**). "Do tone language speakers have perfect pitch?," Proceedings of the Eighth International Conference on Music Perception and Cognition, Evanston, IL, edited by S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, and P. Webster, Causal Productions, Adelaide, Australia, p. 350.

Deutsch, D. (**1991**). "The tritone paradox: An influence of language on music perception," Music Percept. **8**, 335–347.

Deutsch, D. (**2006**). "The enigma of absolute pitch," Acoustics Today **2**, 11–19.

Deutsch, D., Henthorn, T., and Dolson, M. (**2004a**). "Absolute pitch, speech, and tone language: Some experiments and a proposed framework," Music Percept. **21**, 339–356.

Deutsch, D., Henthorn, T., and Dolson, M. (**2004b**). "Speech patterns heard early in life influence later perception of the tritone paradox," Music Percept. **21**, 357–372.

Deutsch, D., Henthorn, T., Marvin, E., and Xu, H. (**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.

Deutsch, D., North, T., and Ray, L. (**1990**). "The tritone paradox: Correlate with the listener's vocal range for speech," Music Percept. **7**, 371–384.

Dolson, M. (**1994**). "The pitch of speech as a function of linguistic community," Music Percept. **11**, 321–331.

Gottfried, T. L. (**2007**). in *Language Experience in Second Language Speech Learning*, edited by O.-S. Bohn, and M. J. Munro (John Benjamins, Amsterdam), pp. 221–237.

Gottfried, T. L., and Riester, D. (**2000**). "Relation of pitch glide perception and mandarin tone identification," J. Acoust. Soc. Am. **108**, 2604.

Gottfried, T. L., Staby, A. M., and Ziemer, C. J. (**2001**). "Musical experience and mandarin tone discrimination and imitation," J. Acoust. Soc. Am. **115**, 2545.

Gottfried, T. L., and Suiter, T. L. (**1997**). "Effects of linguistic experience on the identification of mandarin Chinese vowels and tones," J. Phonetics **25**, 207–231.

Greenberg, S., and Zee, E. (**1979**). "On the perception of contour tones," UCLA Working Papers in Phonetics **45**, 150–164 (http://repositories.cdlib.org/uclaling/wpp/No45/.

Honorof, D. N., and Whalen, D. H. (**2005**). "Perception of pitch location within a speaker's F0 range," J. Acoust. Soc. Am. **117**, 2193–2200.

Johnson, K. A. (**2005**), in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Malden, MA), pp. 363–389.

Ladefoged, P. (**2003**), *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques* (Blackwell, Malden, MA).

Leather, J. (**1983**). "Speaker normalization in perception of lexical tone," J. Phonetics **11**, 373–382.

Lee, C.-Y., Tao, L., and Bond, Z. S. (**2006**). "Native and non-native identification of acoustically modified mandarin tones," J. Acoust. Soc. Am. **120**, 3175.

Lee, C.-Y., Tao, L., and Bond, Z. S. (**2008a**). "Identification of acoustically modified mandarin tones by native listeners," J. Phonetics (in press).

Lee, C.-Y., Tao, L., and Bond, Z. S. (**2008b**). "Speaker variability and context in the identification of fragmented mandarin tones by native and nonnative listeners," J. Phonetics (in press).

Lin, T., and Wang, W. S.-Y. (**1984**). "Shengdiao ganzhi wenti (The issue of tone perception)," Zhongguo Yuyan Xuebao (Bull. Chin. Linguist.) **2**, 59–69.

Liu, S., and Samuel, A. G. (**2004**). "Perception of mandarin lexical tones

when F0 information is neutralized," Lang Speech **47**, 109–138.
Lockhead, G. R., and Byrd, R. (**1981**). "Practically perfect pitch," J. Acoust. Soc. Am. **70**, 387–389.
Mertus, J. A. (**2000**). *The Brown Lab Interactive Speech System* (Brown University), http://www.mertus.org/Bliss/index.html. Last viewed 9/29/08.
Moore, C. B., and Jongman, A. (**1997**). "Speaker normalization in the perception of mandarin Chinese tones," J. Acoust. Soc. Am. **102**, 1864–1877.
Ohio University School of Music Undergraduate Handbook (**2007–2008**). Available at http://www.finearts.ohio.edu/music/gfx/media/pdf/undergradhandbook2007.pdf. Last viewed 8/29/08.
Patel, A. D. (**2008**), *Music, Language, and the Brain* (Oxford University Press, New York).
Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–950.
Saffran, J. R., and Griepentrog, G. J. (**2001**). "Absolute pitch in infant auditory learning: Evidence for developmental reorganization," Dev. Psychol. **37**, 74–85.
Slevc, R., and Miyake, A. (**2006**). "Individual differences in second-language proficiency: Does musical ability matter?" Psychol. Sci. **17**, 675–681.
Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.
Tseng, C.-Y. (**1981**). "An Acoustic Phonetic Study on Tones in Mandarin Chinese," Ph.D. thesis, Brown University.
Whalen, D. H., and Xu, Y. (**1992**). "Information for mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.
Wong, P. C. M., and Diehl, R. L. (**2003**). "Perceptual normalization for inter- and intra-talker variation in Cantonese level tones," J. Speech Lang. Hear. Res. **46**, 413–421.
Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (**2007**). "Musical experience shapes human brainstem encoding of linguistic pitch patterns," Nat. Neurosci. **10**, 420–422.
Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**, 247–258.
Zhou, N., Zhang, W., Lee, C.-Y., and Xu, L. (**2008**). "Lexical tone recognition with an artificial neural network," Ear Hear. **29**, 326–335.

The role of temporal cues in word identification by younger and older adults: Effects of sentence context

Sandra Gordon-Salant and Grace Yeni-Komshian
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

Peter Fitzgibbons
Department of Hearing, Speech, and Language Sciences, Gallaudet University, Washington, DC 20002

(Received 6 August 2007; revised 8 August 2008; accepted 17 August 2008)

Prior investigations, using isolated words as stimuli, have shown that older listeners tend to require longer temporal cues than younger listeners to switch their percept from one word to its phonetically contrasting counterpart. The extent to which this age effect occurs in sentence contexts is investigated in the present study. The hypothesis was that perception of temporal cues differs for words presented in isolation and a sentence context and that this effect may vary between younger and older listeners. Younger and older listeners with normal-hearing and older listeners with hearing loss identified phonetically contrasting word pairs in natural speech continua that varied by a single temporal cue: voice-onset time, vowel duration, transition duration, and silent interval duration. The words were presented in isolation and in sentences. A context effect was shown for most continua, in which listeners required longer temporal cues in sentences than in isolated words. Additionally, older listeners required longer cues at the crossover points than younger listeners for most but not all continua. In general, the findings support the conclusion that older listeners tend to require longer target temporal cues than younger normal-hearing listeners in identifying phonetically contrasting word pairs in isolation and sentence contexts.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2982409]

PACS number(s): 43.71.Lz, 43.71.Es, 43.66.Mk, 43.66.Sr [RLF] Pages: 3249–3260

I. INTRODUCTION

The overall objective of this study was to determine the effect of stimulus context on perception of target temporal cues in speech by younger and older listeners. Although investigations of perception of temporal cues in speech continua have often demonstrated performance differences in younger and older listeners [e.g., see Strouse *et al.* (1998) and Gordon-Salant *et al.* (2006)], these studies have primarily presented isolated words containing the targeted temporal cues. It may be that perception of target cues in speech stimuli presented in isolation may not predict perception of these same cues when embedded in a sentence context.

The assessment of perception of temporal cues in speech by younger and older listeners emerges from a larger body of evidence showing that older human listeners exhibit slowed processing of acoustic signals. This conclusion derives from experiments employing nonspeech signals, as well as those employing speech sounds. For example, many investigations have demonstrated an age effect for detection and discrimination of simple nonspeech sounds presented in isolation. Older listeners show longer duration difference limens (DLs) for discrimination of reference stimuli defined as pure tones or silent intervals (Fitzgibbons and Gordon-Salant, 1994; Bergeson *et al.*, 2001; Grose *et al.*, 2006) and longer gap detection thresholds for gaps inserted between tonal markers or noise bands (He *et al.*, 1999; Snell and Hu, 1999; Lister *et al.*, 2002).

Age-related slowing of perceptual processing is evident also in experiments measuring identification of isolated words, which varied on a single temporal cue along a continuum. In one study, older listeners showed more gradual slopes than younger listeners in identification functions derived from judgments of synthetic speech stimuli varying along a voice-onset time (VOT) continuum (Strouse *et al.*, 1998). In another study, Gordon-Salant *et al.* (2006) used natural speech stimuli and found that older listeners required a longer acoustic cue to shift their perception from one phoneme category to another along a stimulus continuum. For example, in a continuum that varied the formant transition duration to cue a stop-glide distinction, younger listeners required a transition duration of about 29 ms to shift their perception from the stop to the glide, whereas older listeners required a transition duration of about 37 ms to shift their perception from the stop to the glide. Thus, for this one continuum, older listeners required the target cue to be approximately 30% longer than the younger listeners to differentiate the contrasting phoneme with the longer cue from the contrasting phoneme with the shorter cue. The age effect in the study of Gordon-Salant *et al.* (2006) was observed for the silent interval cue for a sibilant-affricate distinction (as in *dish* versus *ditch*), in addition to the transition duration cue for a stop-glide distinction described above (as in *beat* versus *wheat*). However, in this previous study, the age effect was not observed consistently across different speech continua, such as those varying VOT (*buy* versus *pie*) and vowel duration (*wheat* versus *weed*).

The temporal cues discussed above are important for distinguishing certain speech sounds in English as well as in other languages (Lisker and Abramson, 1964; Chen, 1970). The critical values of these temporal cues are, however, lan-

guage specific. As a result, non-native speakers of English often produce speech with temporal characteristics that correspond more to their own language than to English. In other words, the accented speech of non-native speakers contains alterations of numerous acoustic characteristics of English, particularly those that involve the temporal aspects of speech, such as VOT (Flege and Eefting, 1988; MacKay *et al.*, 2000) and vowel duration (Chen, 1970; Fox *et al.*, 1995).

These distinctions between native and accented speech, coupled with the age-related differences in perception of temporal acoustic cues, have potential implications for identifying the source of older listeners' difficulty in understanding accented English. At least one report suggests that older listeners have considerable difficulty in recognizing words and sentences spoken by non-native speakers of English (Burda *et al.*, 2003). One possible reason for older listeners' problems in understanding accented English is difficulty distinguishing key words that are cued by temporal contrasts. The prior investigation of perception of temporal cues in isolated words (Gordon-Salant *et al.*, 2006) revealed age-related changes in perception of two cues (silent interval duration and transition duration), but these are not the temporal cues that are cited as being altered in accented English (i.e., VOT and vowel duration). However, it is possible that perception of VOT and vowel duration in a sentence context may be affected differentially by age. The effect of sentence context on perception of temporally mediated target cues is currently unknown.

In psychoacoustic studies, placement of a nonspeech target stimulus in a sequential context is known to alter perception of the target, particularly by older listeners. Fitzgibbons and Gordon-Salant (1995) measured duration DLs for tones and silent intervals, each with a reference duration of 250 ms, presented in isolation and embedded in a sequence of five 250 ms components. The 250 ms standard component duration was chosen to mimic the average duration of a word. Differential effects of sequential context for younger and older listeners varied with the two types of stimuli (tones and silent intervals). Whereas the young listeners showed a constant DL for tones in isolation and in the sequence, older listeners showed significantly larger DLs for the tones in the five-component sequence compared to the DLs for the same tones measured in isolation. However, for reference silent intervals, both younger and older listeners showed larger DLs in the sequence of five components, but the sequential context effects were still larger for the older listeners. The source of the pronounced age-related deficit for sequential stimuli is unknown but may be associated with age-related decline in global cognitive abilities, coupled with a slowing in component processes (Salthouse, 1996). Deficits in working memory, divided attention, and search capacity, as well as limited processing resources for multiple mental, perceptual tasks have all been reported in older people [e.g., see Baddely (1986), Fisk and Rogers (1991), Tun (1998), and Humes *et al.* (2006)]. It is possible that any one of these attributes of cognitive decline, or a combination of these factors, is implicated in an older listener's reduced ability to make judgments about temporally mediated stimuli in a sequential context. Nevertheless, older listeners are able to make excellent use of contextual meaningful information for understanding speech materials (Pichora-Fuller *et al.*, 1995; Gordon-Salant and Fitzgibbons, 1997), particularly in difficult listening situations. If the nature of the contextual information is semantically limited, however, it may be predicted that age-related differences in perception of temporal cues in speech will be magnified in a sentence context.

The purpose of this study was to investigate the hypothesis that listener perception of target temporal cues to identify words depends on the stimulus context (isolated words versus semantically limited sentences). Placing a cue in a sentence context may influence listener perception of that cue because of the mediating effects of phonemes that precede it, the increased memory load, and/or the increased attention required to locate the cue in a surround. A secondary hypothesis was that older listeners require longer acoustic cues than younger listeners to shift their percept from a phoneme with a shorter reference cue to its contrasting phoneme and that these age-related performance differences may interact with stimulus context. The age-related effects could be manifested primarily in the form of longer temporal cues required by older listeners at crossover points and secondarily in more gradual slopes, consistent with previous studies [e.g., see Strouse *et al.* (1998)]. The interaction effect, if observed, could be seen as a larger difference in performance between younger and older listeners in the sentence context because of increased cognitive demand that may have a greater impact on the older listeners.

A final issue in any examination of age-related performance differences on listening tasks concerns differentiating the effects of age from those attributed to hearing loss, particularly because hearing loss has a high prevalence rate among older people [e.g., see Cruickshanks *et al.* (1998)]. In the present study, the effects of age were investigated through a comparison of performance by older listeners with normal hearing to performance by young noise-masked listeners. The detection thresholds of older listeners with normal hearing are often poorer than those of younger listeners with normal hearing. To accommodate this potential confounding effect, a low-level masking noise was presented to a group of young normal-hearing listeners to shift their thresholds to approximate those of the older normal-hearing listeners. An additional, separate group of young normal-hearing listeners without noise masking served as a control group to validate the continua, as well as to determine the extent to which the noise masking affected performance. Finally, the performance of older hearing-impaired listeners was tested as well to assess the effects of hearing impairment on these tasks. The specific effects of hearing impairment are important to evaluate because previous investigations have shown that hearing loss affects the identification of speech stimuli varying in a single temporal acoustic dimension [e.g., see Gordon-Salant *et al.* (2006)].

II. METHOD

A. Participants

Four groups of listeners ($n=15$/group) participated in these experiments. The first three groups included a control

TABLE I. Pure tone thresholds in dB HL (re: ANSI 2004) in the test ear of the four listener groups (n=15/group). Data shown are group means, with standard deviations in parentheses

	Frequency (Hz)				
Group	250	500	1000	2000	4000
Young normal	7.6(4.9)	7.0(4.9)	5.0(4.2)	3.3(8.2)	1.0(6.6)
Young noise masked	11.0(4.3)	13.3(3.1)	15.3(4.0)	18.7(4.0)	20.3(4.0)
Older normal	15.3(5.7)	15.3(5.3)	14.6(6.3)	17.8(8.0)	18.9(6.2)
Older hearing impaired	25.0(9.6)	24.7(10.3)	27.7(10.7)	40.7(10.0)	51.7(7.7)

group of young listeners (Yng Norm; ages 18–29 years, mean age=22.26) with normal hearing (pure tone air conduction thresholds <20 dB HL re: ANSI (2004), 250–4000 Hz), a group of older listeners (Older Norm; ages 65–79 years, mean age=70.93) with normal hearing, and a group of older listeners (Older Hrg Imp; 66–80 years, mean age=72.86) with mild-to-moderately severe sloping sensorineural hearing loss. Because pure tone detection thresholds are often different between younger and older normal-hearing listeners, spectrally shaped noise masking was presented to a separate fourth group consisting of young normal-hearing listeners (Yng Noise Mask; n=15, ages 18–25 years, mean age=21.5) to equate their masked thresholds to the unmasked thresholds of the older normal-hearing group. This masking noise was added to the stimuli presented to this group during all testing conditions. Comparisons of performance between the young noise-masked and older normal-hearing groups were used to assist in the interpretation of the effects of age (Hargus and Gordon-Salant, 1995; Gordon-Salant *et al.*, 2006). The low-level masking noise was created using COOL EDIT PRO (Version 2, Syntrillium Software) to shift the thresholds of these young listeners to be equivalent (within 5 dB) to the average thresholds at each frequency (250–4000 Hz) of the older group with normal hearing. None of the listeners in the young noise-masked group served in the young normal-hearing control group (unmasked). Mean thresholds of the four listener groups are shown in Table I.

All listeners were required to have good to excellent recognition scores for monosyllabic words (Northwestern University Test No. 6; Tillman and Carhart, 1966) presented under earphones, tympanograms demonstrating normal peak admittance, peak pressure, and equivalent volume (Roup *et al.*, 1998), acoustic reflexes elicited at levels that were within the 90th percentile for individuals with normal hearing or a cochlear hearing loss (Gelfand *et al.*, 1990), and transient-evoked otoacoustic emissions that were present in the listeners with normal hearing and absent in the listeners with hearing impairment. These criteria were intended to ensure that listeners with normal hearing did not have significant peripheral auditory dysfunction and that listeners with hearing loss had a primarily cochlear site of lesion. In addition, American English was the native language of all participants. Finally, all listeners were screened for general cognitive awareness (Short Portable Mental Status Questionnaire, Pfeiffer, 1977) and were required to score at least 9 out of 10 on this test.

B. Stimuli

The stimuli were phonetically contrasting word pairs that varied in a single temporal cue. These words were presented either in isolation or in a sentence context. Eight different speech continua were developed for this purpose. The stimuli on each continuum varied in a single temporal cue in seven equal-interval steps. Three of the continua, consisting of nouns (*buy/pie*, *dish/ditch*, and *beat/wheat*), were used in a previous experiment (Gordon-Salant *et al.*, 2006), and five additional continua, all consisting of names (*Greg/Creig, Sue/Stu, Bill/Will, Alisa/Aliza, and Mr. Rite/Mr. Ride)*, were developed specifically for this experiment. The name continua were implemented in order to permit the placement of the stimulus word at the beginning or end locations in a sentence. In most instances, each temporal cue was examined twice in order to determine whether or not patterns of performance were consistent across different continua varying on the same acoustic cue.

A male talker (age 19 years) with a general American dialect recorded all of the speech stimuli with the names and nouns produced in isolation and in the context of a sentence. The names and nouns spoken in isolation were used to create the tokens for the speech continua, as described below. These tokens were then presented either in isolation or inserted into carrier sentences (replacing the recorded unprocessed word at the beginning or end of a sentence). The carrier sentences were taken from or modeled after the low probability Revised Speech Perception in Noise (R-SPIN) lists (Bilger *et al.*, 1984), which do not provide semantic information to cue the test word. Stimuli were recorded directly onto a laboratory PC using a microphone, a preamplifier, and a 16 bit sound card (Audigy Sound Blaster 2ZS). The name stimuli were recorded in the context of multiple sentences, with the names occurring at or near the beginning of a sentence, as well as at the end of a sentence. The noun stimuli were recorded at the end of different carrier sentences and were usually preceded by the word "the." Altogether, there were six different carrier sentences for the name stimuli and 70 different carrier sentences for the noun stimuli. The numbers of carrier sentences for noun and name stimuli were unequal, reflecting the available corpus of appropriate sentences from the R-SPIN lists. Given the nature of the two-alternative forced choice identification task, the lack of semantic information in the carrier sentences, and the repetition of all of the carrier sentences multiple times throughout the experiment,

it was assumed that the different number of carrier sentences for a given type of stimulus had only minimal influence on results.

The digital recordings were edited using COOL EDIT PRO and WEDW (2008). First, the stimuli in each contrasting word pair recorded in isolation were extracted from the recordings. These isolated stimuli, rather than the stimuli recorded in the sentence contexts, were chosen for processing for two reasons: (1) to facilitate comparison with a previous experiment (Gordon-Salant *et al.*, 2006) that also used stimuli recorded in isolation and (2) to permit a direct comparison of perception of the exact same words presented in isolation and in sentence contexts. The words in each pair were analyzed acoustically and modified to create a continuum of seven stimuli in which the single temporal cue of interest was varied in equal steps. It should be noted that in natural speech, more than one acoustic cue may vary between two phonetically contrasting words. Nevertheless, the primary acoustic cue that contributes to the perception of these stimuli is the target temporal cue modified in each of the continua. Four acoustic cues were sampled in a total of eight separate continua: VOT to signal stop-consonant voicing, silent interval duration to distinguish either fricative versus affricate (/ʃ/ versus /tʃ/) or the fricative /s/ from the fricative-stop cluster /st/, formant transition duration as a cue for stop versus semivowel, and vowel duration as a cue for voicing in postvocalic consonants.

A complete description of the creation of the noun continua (*buy/pie*, *dish/ditch*, and *beat/wheat*) can be found in a previous publication (Gordon-Salant *et al.*, 2006). They will be described only briefly below. Stimulus development for the proper name stimuli was comparable to that used for the noun stimuli. Detailed descriptions of the creation of these name stimuli (*Greg/Creig*, *Sue/Stu*, *Bill/Will*, *Alisa/Aliza*, and *Rite/Ride*) are also presented.

Voice-onset time. VOT was altered in the *buy/pie* continuum by varying the duration of the aspiration in the natural token, *pie*, from 0 ms (*buy*) to 60 ms (*pie*) in 10 ms steps. The *Greg/Creig* continuum also manipulated VOT. The endpoint stimulus, *Creig*, was modified from the original recording by first removing its burst and 100 ms of aspiration, and then inserting a burst from a natural *Greg*. Stimuli on the seven-point continuum were created by adjusting the amount of aspiration in 10 ms steps, from 20 ms (*Greg*) to 80 ms (*Creig*).

Silent interval duration. The target cue in the *dish/ditch* continuum was the silent interval duration between the initial consonant-vowel and the final consonant, /ʃ/, from 0 ms (*dish*) to 60 ms (*ditch*) in 10 ms steps. The target cue for the *Sue/Stu* continuum was the silent interval duration between /s/ and the final vowel. The endpoint stimulus, *Sue*, was a hybrid; it was created by modifying the original *Stu* stimulus as follows. First, the /s/ from *Stu* was removed and replaced with the natural /s/ from *Sue*. Next, the burst from *Stu* was attenuated by 12 dB, and the 92 ms silent period in *Stu* was removed. This produced the endpoint stimulus perceived as *Sue*, with a 0 ms silent interval. The remaining stimuli on the continuum were created by inserting silent segments, in 5 ms steps, following the fricative /s/ in the endpoint *Sue* stimulus. The endpoint stimulus, *Stu*, had a silent interval of 30 ms.

Transition duration. For the *beat/wheat* continuum, the transition duration of the initial consonant was varied from 7 ms (*beat*) to 51 ms (*wheat*) in 7–8 ms steps (i.e., a single glottal pulse). The *Bill/Will* continuum was created by reducing the natural transition duration of the original stimulus, *Will*, down to 60 ms from 160 ms. This endpoint stimulus was perceived as *Will*. Single pulses of the glide (approximately 10 ms/pulse) were removed to create the remaining stimuli on the continuum, with the endpoint, *Bill*, having a transition duration of 0 ms. The endpoint stimuli *Bill* and *Beat* did not contain a burst, although they were perceived as the intended phoneme (/b/).

Vowel duration. The target cue in the *Alisa/Aliza* continuum was vowel duration as a cue for voicing contrasts in post-vocalic fricatives (Denes, 1955). The endpoint stimulus, *Aliza*, was the natural stimulus whose vowel duration was 356 ms. The duration of the vowel, /i/, was cut from this endpoint stimulus by two pulses (approximately 19–20 ms) for each successive stimulus, such that the vowel of the endpoint stimulus, *Alisa*, was 242 ms. The *Rite/Ride* continuum varied vowel duration as a cue to final consonant voicing. (Note: these stimuli were preceded by Mr. or Ms., as in *Mr. Ride*, to make these stimuli proper nouns.) The endpoint stimulus, *Ride*, was a hybrid of the original recording of *Ride*, with the final consonant release of *Ride* removed and replaced with the release of *Rite*. The duration of the consonant, /r/, and the closure duration for the final consonant were held constant, but the steady-state vowel portion was modified to produce the different stimuli in this continuum. The steady-state vowel portion in the endpoint stimulus, *Ride*, was 294 ms. Additional stimuli in the continuum were created by removing two pulses per step (approximately 19–20 ms) from the steady-state portion of the vowel, such that the vowel duration of the endpoint stimulus, *Rite*, was 174 ms.

The isolated word stimuli from each continuum were equated in rms level. As noted earlier, carrier sentences were recorded into which the processed stimuli from a continuum were inserted. The noun stimuli were inserted at the end of a sentence. There were 70 carrier sentences used for the noun stimuli in each continuum (seven steps/continuum and ten repetitions of each token). The name stimuli were inserted at the beginning of a carrier sentence ("*Sue* heard Tom called about the coach.") and at the end of another carrier sentence ("Ruth could not have known about *Sue*."). This pattern was used for the rest of the contrasting name stimuli. The natural pauses in the original sentences were preserved when creating these sentences. The sentence stimuli across all continua were analyzed and equated in rms level. Two separate calibration tones were created to be equivalent to the rms levels of the isolated word stimuli and the sentence stimuli, respectively. The isolated word and sentence stimuli, as well as the calibration tones, were stored in separate waveform files on the PC.

C. Procedure

Each of the eight speech continua were presented both as isolated words (eight conditions corresponding to the

three noun and five name continua) and as words in a sentence (three conditions corresponding to the noun stimuli placed at the end of sentences and ten conditions corresponding to the five name stimuli, which were placed at the beginning and at the end of sentences) for a total of 21 (8+3+10) conditions. Half of the participants were tested on the isolated words first, and half of the participants were tested on the sentences first. The order of presentation of the specific speech continua was randomized separately for the isolated words and the sentences, and this order also varied across participants.

During testing, the listener was seated in a double-walled sound attenuating chamber. ECOS2 software (Avaaz Innovations) controlled stimulus presentation and timing, presented visual displays to the listener, and tracked the listener's responses. For each condition, the seven stimuli on the target continuum were presented ten times in randomized order, with each randomization occurring within a block of the seven stimuli (i.e., ten randomized blocks were presented). Speech stimuli were played out of a sound card (Audigy Sound Blaster 2ZS), amplified (Crown D-75A), mixed (Coulbourn S-24 audiomixer amplifier), and delivered to an insert earphone (Etymotic ER-3A). The signal level was 85 dB SPL. For the young noise-masked listeners, a CD of the masking noise was played out of a CD player (Tascam model CD-RW402), amplified (Crown D-75A), attenuated (Hewlett-Packard 350D), and mixed with the target speech signal (Coulbourn audiomixer amplifier). The level of the spectrally shaped noise was adjusted to produce 50 dB SPL in the insert earphone, as measured in a 2 cm^3 coupler. The test ear was the ear with better hearing sensitivity for the hearing-impaired listeners and was randomized for the normal-hearing listeners.

Following the stimulus presentation, a computer monitor presented a text display of the two possible response choices, corresponding to the two endpoint stimuli on the continuum (e.g., *buy* and *pie*). The listener's task was to select the stimulus perceived using a computer mouse click. Listener responses were self-paced. Prior to commencing each condition, instructions were given orally and in a visual display on the computer monitor. Before the experimental conditions began, listeners were given practice using another speech continuum that was created exclusively for practice purposes.

The entire experiment, including preliminary audiometric testing and experimental testing, was concluded in approximately 6 h, scheduled in several 2 h visits. Listeners were provided frequent breaks to minimize fatigue and were reimbursed for their participation in the experiment. The protocol for this experiment was approved by the University of Maryland Institutional Review Board for Human Subjects Research.

D. Analysis

Two values were calculated from each participant's identification function for each continuum: crossover point and slope. In addition, the endpoint accuracy was calculated for the stimuli presented in isolation, as an indicator of how well the listeners identified the exemplar stimuli (stimuli 1 and 7 on each continuum); lower accuracy indicates that the listeners had a relatively unstable perception of the target stimulus (Munson and Nelson, 2005).

The slope and 50% crossover point of each individual's identification function for each stimulus continuum were estimated using a fitting procedure described by Wichman and Hill (2001a, 2001b) and the accompanying software (PSIGNIFIT, available from http://bootstrap-software.org/psignifit/). The logistic option was used to characterize the shape of the underlying psychometric distribution of the measured identification functions. The estimated crossover point for each function represents the stimulus value at which a listener changes his or her perception from one stimulus to the contrasting stimulus on the continuum. The slope of the function indicates how abrupt and consistent the shift is in perception from one phoneme category to another. Steep slopes may suggest that the stimuli on the continuum were perceived with certainty, whereas a gradual slope reflects the uncertainty in perception of the stimuli encompassed by the slope.

Some participants were random in their responses. Data from these participants were removed from the analysis for a particular condition if the identification scores of both endpoint stimuli did not exceed 50%, suggesting that the listener did not perceive the two endpoint stimuli as distinct. In these cases, a crossover point and slope are not meaningful. A subset of these participants perceived all of the stimuli along the continuum as one of the endpoint stimuli (e.g., flat slope). The PSIGNIFIT program could not derive crossover points and slopes for these flat responses. In all, there were several participants who showed random performance for a particular condition, although the specific participants who exhibited random identification performance on one condition were not necessarily those who showed random performance on other conditions. Participants with random endpoint accuracy on a particular condition came from all listener groups. The proportion of data dropped from each listener group, across all 21 conditions, was 5.7% for young normal controls, 6.0% for young noise-masked listeners, 9.2% for older normal-hearing listeners, and 10.8% for older hearing-impaired listeners. The relatively small proportion of responses that were dropped from the total data set of each group suggests that these stimuli were generally perceived as the intended words and hence are reasonably natural and valid. Goodness of fit for each individual function was determined by calculating the correlation coefficients between the PSIGNIFIT predicted data points and the observed data points. The means of the correlation coefficients for the functions associated with the retained data for isolated word and sentence stimuli were 0.98 and 0.97 (R^2=0.96 and 0.94), respectively, while those associated with the rejected data for isolated word and sentence stimuli were 0.42 and 0.56 (R^2=0.18 and 0.31), respectively.

III. RESULTS AND DISCUSSION

Participants in the young normal-hearing group served as the control group to verify the validity of the stimuli. Figure 1 plots the percentage of "buy" judgments as a func-

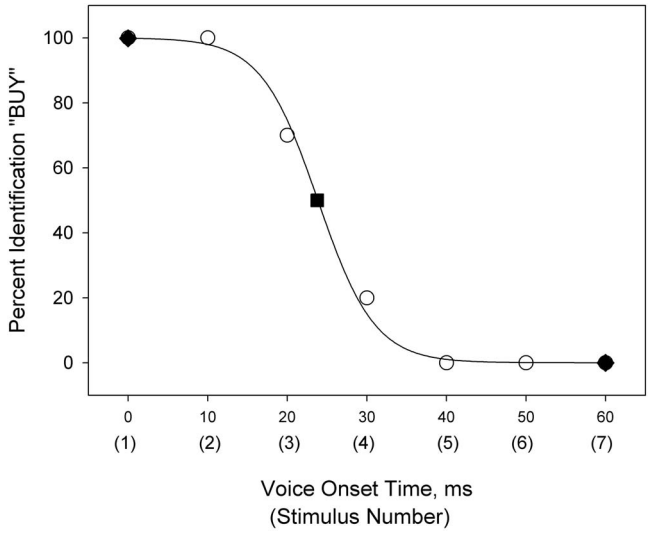


FIG. 1. Sample identification function for the *buy/pie* continuum of one listener from the young normal-hearing (Yng Norm control group). The *x*-axis is labeled with the actual VOT of stimuli on the continuum as well as the stimulus number corresponding to each VOT value (in parentheses). Open circles are the raw data points, the line is the PSIGNIFIT fitted function to these data, the filled square is the PSIGNIFIT derived crossover point, and the filled diamonds are the PSIGNIFIT derived endpoints.

tion of VOT for a representative young normal-hearing participant. Actual data points are shown (open circles) as well as the PSIGNIFIT fitted function for these data and the derived crossover point (filled square) and endpoints (filled diamonds). The figure shows that this listener categorized the shortest VOT (stimulus 1) as "buy" and the longest VOT (stimulus 7) as "pie." The performance of this listener on this stimulus pair was typical of the performance of all of the listeners on most of the stimulus pairs, indicating that participants heard the endpoint stimuli as intended. Typically, the crossover point for all listeners occurred in the region defined by stimuli 3, 4, and 5. The PSIGNIFIT curve also demonstrates the accuracy of the fitted function for data of this form.

A. Endpoint accuracy for isolated stimuli

Percent identification scores were analyzed for the two endpoint stimuli presented in isolation because these stimuli served as the baseline stimuli. Analyses were conducted separately for stimuli from each of the eight continua, following arc-sine transformations of the identification scores. In general, the level of performance of all groups was high for these endpoint stimuli on each continuum (range $=87\%-100\%$; mean$=97.01\%$). This is not unexpected, given that responses on continua demonstrating random performance were eliminated from the analyses. Separate analyses of variance (ANOVAs) were conducted for each continuum using a split-plot factorial design with one between-subjects factor (group: four levels) and one within-subjects factors (stimulus: two endpoints). In this and all subsequent analyses, interaction effects were examined with simple main effect analyses and multiple comparison testing (Bonferroni method) as appropriate.

Overall, effects of group were few in these analyses, indicating that the four listener groups were reasonably well equated for recognition of the endpoint stimuli in each continuum presented in isolation (the baseline condition). A significant main effect of group $[F(3,49)=3.60,\ p<0.05]$ was observed for one continuum in the domain of vowel duration (*Alisa/Aliza)*, but not in the other continuum for vowel duration (*Rite/Ride)*. The significant effect reflected higher performance by the two younger groups compared to the older normal-hearing group. A significant stimulus by group interaction was observed for one VOT continuum (*buy/pie*) $[F(3,54)=3.57,\ p<0.05]$ but not for the other VOT continuum (*Greg/Creig*). The source of the significant interaction was higher recognition scores for *pie* (100%) than for *buy* (96%) for the young noise-masked listeners only.

The only significant main effects of stimulus were observed for one VOT continuum [*Greg/Creig*: $F(1,51)=11.28,\ p<0.01$] and one continuum varying formant transition duration [*Bill/Will*: $F(1,51)=8.45,\ p<0.01$]. Significant effects were not observed for the other VOT continuum (*buy/pie*) or for the other formant transition duration continuum (*beat/wheat*). Listeners, on average, showed higher recognition scores for the endpoint stimulus *Greg* (97%) compared to the endpoint stimulus *Creig* (92%). Although the endpoint stimulus *Will* was identified with higher accuracy than *Bill*, it is noteworthy that the mean performance scores exceeded 90% for this continuum. In summary, all target stimuli were perceived highly accurately in the baseline condition in all listener groups.

B. Crossover points

Each continuum was structured such that the endpoint stimulus with the shorter acoustic cue served as the reference stimulus. Crossover points that were closer to this reference stimulus were expressed as shorter values (in milliseconds) than those that were closer to the alternate endpoint. Thus, a longer crossover point would indicate that a given listener (or group) requires a longer acoustic cue, e.g., VOT, to shift their percept from voiced to voiceless.

1. Voice-onset time

The mean crossover points for the four listener groups for the two continua that varied VOT (*buy/pie, Greg/Creig*), presented in isolation and in the sentence contexts, are shown in Fig. 2. For the *buy/pie* VOT continuum, it is apparent that these crossover points vary considerably for stimuli presented in isolation and in the sentence context and between listener groups. For a few participants, the estimated crossover point was beyond the range of stimulus values; nevertheless, data of these participants were included in the analysis. This outcome is due to the Psignifit fitting function estimates in cases where the listener's identification responses were predominantly for one stimulus on the continuum, with an abrupt shift to the percept of the contrasting phoneme for the other endpoint stimulus. ANOVAs conducted on the crossover points revealed significant main effects of context $[F(1,53)=65.34,\ p<0.01]$ and group $[F(3,53)=23.46,\ p<0.01]$ and a significant interaction be-

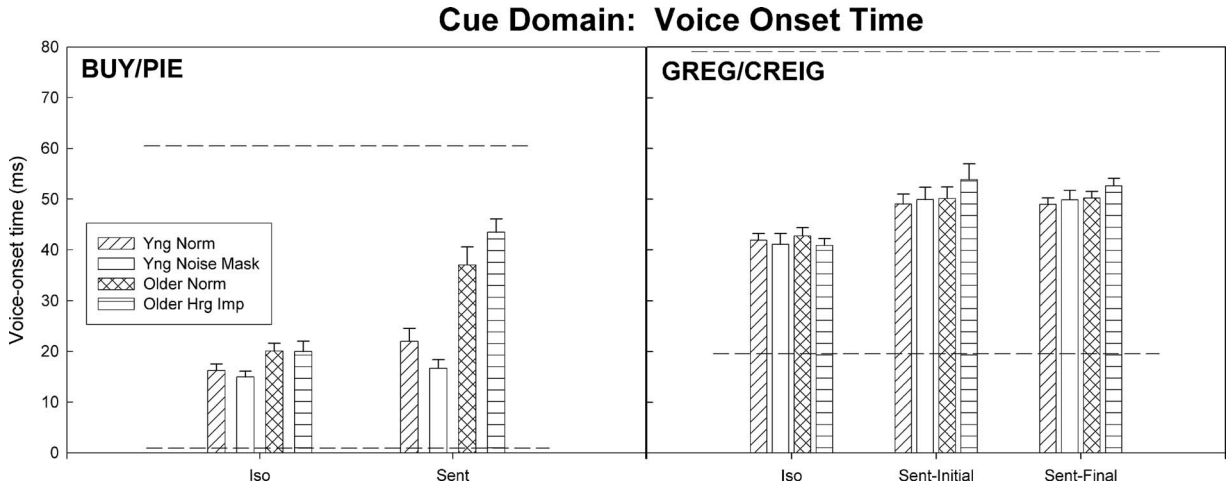


FIG. 2. Mean crossover points (in ms) for the VOT cue for the four listener groups measured in two continua (*buy/pie* and *Greg/Creig*) presented as isolated words (Iso) and in a sentence context (Sent) for the *buy/pie* continuum, and as isolated words (Iso) and in a sentence-initial position (Sent-Initial) and in a sentence-final position (Sent-Final) for the *Greg/Creig* continuum. Error bars show 1 standard error of the mean. The dashed horizontal lines indicate the minimum and maximum VOT values for the endpoint stimuli on each continuum.

tween the two [$F(3,53)=11.80$, $p<0.01$]. *Post hoc* testing showed that the group effect was significant for the sentence context only ($p<0.01$), with the two older groups exhibiting longer crossover points than the two younger groups.

The ANOVA on the crossover points for the *Greg/Creig* continuum also assessed the main effects of context and group, but the context effect had three levels (the design for all of the "name" continua): words in isolation, words at the initial position of sentences, and words at the final position of sentences. A significant main effect of context [$F(2,76)=59.57$, $p<0.01$] was observed. The main effect of group and the interaction effect were not significant. A *post hoc* analysis of the context effect indicated that crossover points were longer for the words presented in sentences (either at the beginning or at the end of the sentence) than for isolated words. There were no significant differences in crossover points for words occurring in the two-sentence locations.

One hypothesis was that age-related differences in perception of target temporal cues would be altered when the target was presented in the context of a sentence. Thus, a group by context interaction effect was predicted. This predicted interaction effect was observed for the crossover points in the *buy/pie* continuum. For this continuum, there were no group differences in the crossover points for the words presented in isolation, consistent with previous findings (Gordon-Salant *et al.*, 2006). However, for words presented in the sentence context, there was a significant group effect in which the crossover points for the two older groups were longer than those for the two younger groups. This result indicates that older listeners required a longer VOT to perceive an ambiguous token on the *buy/pie* continuum as *pie* in a sentence context compared to younger listeners. This is similar to psychoacoustic findings indicating that older listeners show longer DLs for tones presented in a sequence compared to tones presented in isolation, but younger listeners show stable DLs in these two conditions (Fitzgibbons and Gordon-Salant, 1995).

For the *Greg/Creig* continuum, the only significant effect was for context, in which all listener groups required longer VOTs to switch their perception from the voiced *Greg* to the voiceless *Creig* in the sentence context. The divergent findings across the two VOT continua may be related to differences in the specific nature of the VOT cue in these two sets of stimuli. Although both continua varied the duration of the aspiration following a burst as a cue for initial stop voicing, the endpoint stimulus *buy* had a VOT of 0 ms (i.e., no aspiration), whereas the endpoint stimulus *Greg* had a VOT of 20 ms (i.e., a 20 ms aspiration). This is characteristic of VOT values for word initial velar stops reported by Lisker and Abramson (1964). For listeners to shift their percept from *buy* to *pie* may therefore require the detection of the presence of a brief segment of aspiration, whereas for listeners to shift their percept from *Greg* to *Creig* may require the discrimination of a longer segment of aspiration from a shorter one. Clearly, the young listeners (with normal hearing and noise masking) were good at detecting the presence of the aspiration in the *buy/pie* continuum, regardless of the stimulus context (isolation or in sentences), and the older listeners were also reasonably good at this task for the isolated word stimuli. When the stimuli were embedded in the sentence context (final position), the detection of the aspiration segment was more difficult for the older listeners and may be analogous to age effects observed in temporal resolution tasks when the relative location of a brief interruption in a sound is shifted toward the temporal boundary of the stimulus (He *et al.*, 1999). In contrast, the *Greg/Creig* continuum may be more similar to a duration discrimination task for stimulus intervals defined by aspiration. The sentence context effects observed with the *Greg/Creig* continuum appear to parallel the sequential context effects observed in psychoacoustic investigations of duration discrimination of silent intervals by younger and older listeners (Fitzgibbons and Gordon-Salant, 1995), in which both listener groups showed larger DLs for intervals presented in sequences than in isolation. These observations underscore the importance of interpreting context effects in relation to the specific temporal cue under investigation. However, these effects need to be

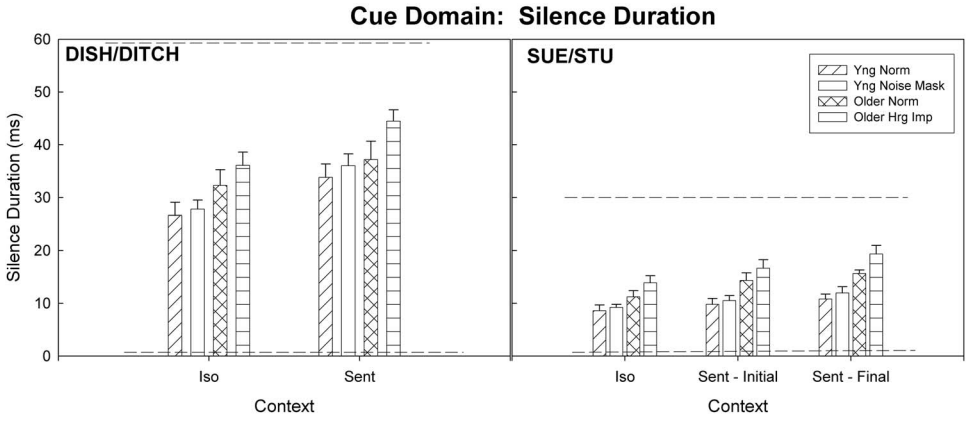


FIG. 3. Mean crossover points (in ms) for the silence duration cue for the four listener groups measured in two continua (*dish/ditch* and *Sue/Stu*) presented as isolated words (Iso) and in a sentence context (Sent) for the *dish/ditch* continuum, and as isolated words (Iso) and in a sentence-initial position (Sent-Initial) and in a sentence-final position (Sent-Final) for the *Sue/Stu* continuum. Error bars show 1 standard error of the mean. The dashed horizontal lines indicate the minimum and maximum silence duration values for the endpoint stimuli on each continuum.

replicated with different tokens of the same target cue, spoken by either the same talker or different talkers.

2. Silent interval

The two continua that varied the duration of a silent interval signaled a shift from a fricative to an affricate (*dish/ditch*) or a shift from a fricative to a consonant cluster (*Sue/Stu*). Mean crossover points derived for these two continua in isolation and sentence contexts are presented in Fig. 3. For the *dish/ditch* continuum, the analysis of the data indicated significant main effects of context $[F(1,54)=34.02,\ p<0.01]$ and group $[F(3,54)=4.00,\ p<0.05]$ but no significant interaction. The main effect of context reflected the longer crossover points observed for the sentence context than for isolated words. Multiple comparison testing revealed that the older hearing-impaired group had longer crossover points than the two young listener groups.

The target cue in the *Sue/Stu* continuum (Fig. 3) was a silent interval duration preceding the final vowel. The ANOVA revealed significant main effects of context $[F(2,92)=27.01,\ p<0.01]$ and group $[F(3,46)=10.68,\ p<0.01]$ and no significant interaction. All listeners exhibited longer crossover points for words in sentences compared to isolated words and longer crossover points for words occurring at the end of the sentence than at the beginning of the sentence. The group effect was attributed to longer crossover points for the older hearing-impaired group compared to the two younger groups. Additionally, the older listeners with normal hearing showed longer crossover points than young listeners with normal hearing.

Overall, similar significant main effects of context and group were observed for the two continua that varied the silence duration cue. The context effect for the *dish/ditch* continuum was attributed to longer crossover points for words in sentences than in isolation. In the *Sue/Stu* continuum, all listeners showed longer crossover points for words in the final sentence position than words in isolation. The final sentence position for these name stimuli is comparable to the location of the target word in sentences for the noun stimuli. Thus, the context effects for these two continua were quite similar.

The group effects in these two continua were also highly comparable. *Post hoc* analyses revealed that for both the *dish/ditch* and *Sue/Stu* continua, the older hearing-impaired listeners demonstrated longer crossover points than the two groups of younger listeners. Both of these continua varied the silent interval duration as a cue for the distinction /ʃ/ versus /tʃ/ or the distinction /s/ versus /st/ ; these consonants are classified as sibilants. This effect of hearing impairment was observed previously for the *dish/ditch* continuum presented as isolated words (Gordon-Salant *et al.*, 2006). It is noteworthy that the range in duration of the silent intervals was twice as long for the *dish/ditch* continuum (0–60 ms) as for the *Sue/Stu* continuum (0–30 ms), yet the same group effect was observed. The listeners in the older hearing-impaired group had moderate hearing losses in the high frequencies, and it is likely that the high frequency cues for the sibilant consonants in the target words were less audible in this group compared to the two young groups with normal hearing, even at the high presentation level used in the study. As a result, hearing-impaired listeners may have experienced difficulty in coding the onset and offset of sibilance and in detecting the relatively brief silent interval that is contiguous with these sibilant consonants. Several studies have shown that hearing-impaired listeners have larger gap detection thresholds than normal-hearing listeners when the gap is bounded by a broadband noise burst (Irwin *et al.*, 1981; Florentine and Buus, 1984) or when the gap is bounded by two different markers, such as a broadband noise burst and a tone (Grose *et al.*, 1989; Lister *et al.*, 2000). Thus, reduced ability to detect a silent gap between two markers may contribute to the hearing-impaired listeners' use of the silent interval cue, although this was not assessed directly in the present investigation.

3. Transition duration

The temporal cue for the *beat/wheat* and *Bill/Will* continua was transition duration; the mean crossover points of the four listener groups are shown in Fig. 4. Results of statistical analysis for the *beat/wheat* continuum revealed significant main effects of context $[F(1,44)=48.86,\ p<0.01]$ and group $[F(3,44)=6.89,\ p<0.01]$ and no significant interaction. For this continuum, crossover points were longer for isolated words than for the stimuli in the sentence context. The group effect was attributed to longer crossover points for the two older groups compared to the young normal-hearing group.

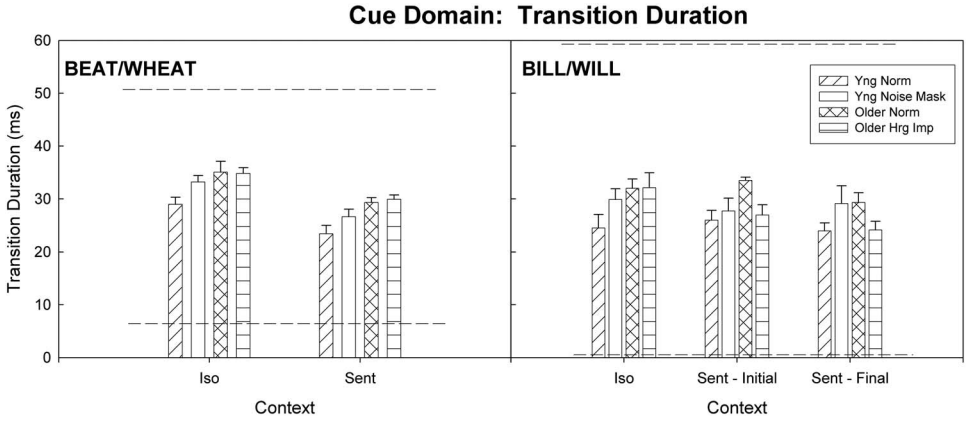


FIG. 4. Mean crossover points (in ms) for the transition duration cue for the four listener groups measured in two continua (*beat/wheat* and *Bill/Will*) presented as isolated words (Iso) and in a sentence context (Sent) for the *beat/wheat* continuum, and as isolated words (Iso) and in a sentence-initial position (Sent-Initial) and in a sentence-final position (Sent-Final) for the *Bill/Will* continuum. Error bars show 1 standard error of the mean. The dashed horizontal lines indicate the minimum and maximum transition duration values for the endpoint stimuli on each continuum.

Somewhat similar results were observed for the *Bill/Will* continuum. Statistical analyses revealed a significant main effect of context only [$F(2,90)=5.79$, $p<0.01$]. Multiple comparison testing indicated that crossover points were longer for isolated words than for words in the sentence-final position.

Taken together, the two continua that varied transition duration as a cue to the consonant manner (stop/semivowel) exhibited context effects, based on an analysis of the crossover points, that were highly similar and quite different from those observed for the other continua. Specifically, listeners from all groups showed longer crossover points for the isolated words than for the words in the final position of a sentence. One possible explanation for this observation is that listeners relied on the rapid transition to cue the initial stop, /b/, since these stimuli did not contain a burst. For words presented in isolation, listeners perceived transition durations of 0–20 ms (on average) as /b/, which is consistent with the expected duration of this cue for this phoneme (Liberman *et al.*, 1956; Lisker and Abramson, 1964). When the target noun was embedded in the sentence-final position, it was usually preceded by the vowel /ə/. This may have produced a continuous transition from the vowel /ə/ to the onset of the target stimulus, which in turn may have enhanced the percept of the semivowel /w/ with shorter transition durations. Listeners required an extremely brief transition to signal the onset of the target word as *beat* following this vowel. The same argument does not apply for the name continuum, however, because the phoneme preceding the target was a voiceless stop.

Group effects were observed for the *beat/wheat* continuum. Younger listeners showed shorter crossover points than the two older listener groups, and this finding is similar to a previous report for the *beat/wheat* continuum presented as isolated words (Gordon-Salant *et al.*, 2006). For the *Bill/Will* continuum, there was considerable variability in the crossover points, which likely precluded the observation of any age effects. It is not clear why there was a large variability in the crossover points for the young listeners with normal hearing, especially in light of the fact that the endpoint accuracy for these stimuli was quite high for all listener groups (>94%).

4. Vowel duration

The last two continua, *Alisa/Aliza* and *Rite/Ride*, varied the vowel duration as the target cue for postvocalic consonant voicing. An inspection of the crossover distribution for *Alisa/Aliza* (Fig. 5) suggests that younger listeners perform differently than the older listeners, and the ANOVA confirmed this impression. A significant main effect was observed for listener group [$F(3,42)=16.18$, $p<0.01$]. *Post hoc* testing revealed that the two older groups had longer crossover points than the two younger groups. None of the other effects were significant. For the *Rite/Ride* continuum (Fig. 5), the results showed a significant main effect of context [$F(2,108)=5.37$, $p<0.01$] and a significant main effect of group [$F(3,54)=4.24$, $p<0.01$], with no significant interaction between these two variables. A *post hoc* analysis of the group effect showed a significant difference in mean per-

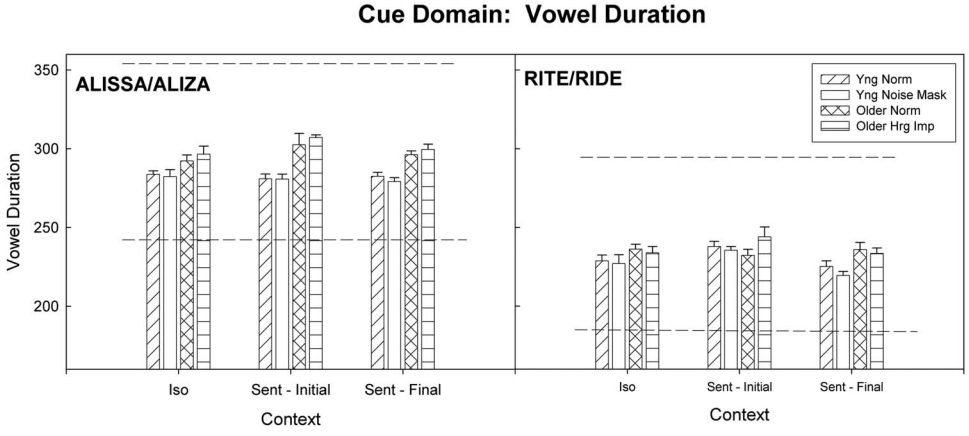


FIG. 5. Mean crossover points (in ms) for the vowel duration cue for the four listener groups measured in two continua (*Alisa/Aliza* and *Rite/Ride*) presented as isolated words (Iso) and in a sentence-initial position (Sent-Initial) and in a sentence-final position (Sent-Final). Error bars show 1 standard error of the mean. The dashed horizontal lines indicate the minimum and maximum vowel duration values for the endpoint stimuli on each continuum.

formance between the young noise-masked listeners and the two older listener groups, which was attributed to longer crossover points for the older groups. Crossover points for words in the sentence-initial position were overall longer than for words in the sentence-final position as well as for words in isolation.

Observed group effects were similar for these two continua. Analyses of crossover points for the *Alisa/Aliza* continuum revealed that the two groups of younger listeners demonstrated shorter crossover points than the two groups of older listeners. In the *Rite/Ride* continuum, the young noise-masked listeners also showed shorter crossover points than the two groups of older listeners. This group difference in crossover points was found consistently for stimuli presented in isolation and in the sentence context for both continua. For both continua, then, older listeners had longer crossover points than younger listeners. Because vowel duration is a well-established cue for postvocalic voicing in stops (Peterson and Lehiste, 1960) and fricatives (Denes, 1955), the current findings indicate that the older listeners required longer vowel durations than younger listeners to alter their percept of postvocalic consonantal voicing.

A context effect was also shown for the *Rite/Ride* continuum. All listeners showed longer crossover points for words presented in the sentence-initial position than for words presented either in isolation or in the sentence-final position. This is the only continuum in which this pattern of context effects was observed. The presentation of the target item in the initial position of the sentence was somewhat different in this continuum than in all other continua because the target was embedded in the phrase "Mr. Rite" or "Mr. Ride." As a result, the target cue was actually embedded in the second word in the sentence, rather than the first word. This may have made the target cue less perceptible or less accessible to the listeners in this location compared to the words in isolation or in the sentence-final position.

C. Slopes

Similar statistical analyses to those reported above were conducted for the slope values derived from the PSIGNIFIT analysis of each listener's identification function for each condition. In general, the analyses of the slope values did not reveal significant effects. Of the 24 effects analyzed (8 continua $\times$ 3 effects/continuum), only three were significant. The effects of context only were significant for one VOT continuum [*buy/pie;* $F(1,54)=20.46$, $p<0.01$] and one transition duration continuum [*beat/wheat* $F(1,44)=12.15$, $p<0.01$]. Slope values were higher for the isolated word stimuli compared to the words in sentences for both of these continua. It should be noted that significant effects of context on slope values were not observed for the other VOT continuum (*Greg/Creig*) or the other transition duration continuum (*Bill/Will*). A significant main effect of group was observed for one continuum varying silent interval duration [*Sue/Stu*: $F(3,46)=4.53$, $p<0.01$]. Multiple comparison testing showed that the young noise-masked listeners exhibited steeper slopes than the two older listener groups on this continuum. Group effects were not observed for the other continuum varying silent interval duration (*dish/ditch*). In summary, the identification function slopes varied minimally between groups and contexts for the different continua.

IV. GENERAL DISCUSSION

The principal hypothesis concerning differential performance by all listener groups in the two stimulus contexts was observed for many temporal cues. In most cases, the sentence context effect (when present) resulted in longer crossover points for sentences than for individual words. The older listener groups also perceived the target temporal cues differently from the younger listener groups for many stimulus continua.

Age-related differences in the perception of temporal cues in natural speech continua were observed in this study in most, but not all, continua. In all cases where age effects were observed, older listeners required a longer temporal cue to alter their percept from the reference stimulus (that containing the shorter cue) to the alternate stimulus. The strongest age-related effects were observed in the two continua that varied the vowel duration as a cue to postvocalic voicing (*Alisa/Aliza* and *Rite/Ride*), and these age effects were seen for all stimulus contexts. Age effects were also observed in one continuum that varied the transition duration (*beat/wheat*) for stimuli presented in both word and sentence contexts. A modest age effect was observed for the *Sue/Stu* continuum as well. Finally, strong age effects were observed for one VOT continuum (*buy/pie*) in the sentence context only. These observations are consistent with the notion that for many stimuli, older listeners have difficulty processing relatively brief changes in duration compared to younger listeners. The present findings suggest, however, that the problems older people have in the perception of brief target temporal cues in speech are cue specific. The duration of temporal cues in natural speech is highly variable and dependent on the talker, the stimulus context, and the specific target phonemes. The current results show that a different pattern of age-related effects can emerge for the same type of temporal cue when it appears in different phonemes. This may be one contributing factor to the common observation of wide variability in speech recognition performance of older listeners.

These findings may have implications for the perception of accented speech by elderly listeners. For example, stimuli in the middle of a speech continuum may be similar to those produced in accented English. A native Spanish speaker of English will produce the intended word, *pie*, with a shorter VOT than a native English speaker, resulting in the perception of this word as *buy* by a native English listener. Because older listeners require longer VOTs than younger listeners to perceive *pie* in sentence contexts, it is even more likely that they will misperceive the intended *pie* produced by native Spanish speakers. A similar phenomenon may occur as a result of altered vowel duration in accented speech. Nonnative speakers of English often produce shorter vowels than native speakers, resulting in a greatly reduced distinction in vowel duration as a cue to postvocalic consonant voicing (Chen, 1970). This could result in the perception of a postvocalic consonant as voiceless rather than voiced, particularly

by older listeners who require longer vowel durations in the region of uncertainty to shift their perception to the voiced cognate. In the current experiment, the ambiguous words in the middle of the continuum were presented in the context of unaccented English. At present, it is unknown if embedding these same words in the context of accented English would alter the listener perception of the ambiguous target words. It is possible that listeners would have less difficulty because of perceptual normalization, but, alternatively, listeners may have more difficulty because contextual cues may not be well understood. At least one report (Burda *et al.*, 2003) has shown that older listeners recognize accented English more poorly than young adult listeners. The current findings are useful for explaining some possible underlying mechanisms for this observation.

Age-related differences in identification function slopes were not observed for most continua. It was anticipated that older listeners, who often exhibit poorer temporal resolution than younger listeners, would show more gradual identification function slopes on speech continua that varied temporal cues compared to younger listeners. For example, Strouse *et al.* (1998) reported significantly different slope values between younger and older listener groups for the perception of a synthetic speech continuum varying from /ba/ to /pa/. There are some procedural differences between the study of Strouse *et al.* (1998) and the present study that might account for differing outcomes regarding slope values. First, the stimuli in the present study were natural speech tokens, rather than synthetic speech tokens as used in the study of Strouse *et al.* (1998). With real speech, there are multiple cues in addition to the temporal cue that was manipulated in the present study. Listeners may have made use of these cues. Second, the current investigation used the PSIGNIFIT analysis (a maximum likelihood method with the selection of a logistic form of the psychometric function) to estimate the crossover points and slopes from the identification data, whereas the study of Strouse *et al.* (1998) used a linear regression analysis to derive crossover points and slopes. These methods for estimating the slopes are quite different and possibly contributed to different outcomes. Nevertheless, the current analysis of the data revealed clear differences in crossover durations required by younger and older listeners in many of the speech continua. Taken together, these findings suggest that for many of the temporally mediated continua, the entire function appears to be shifted to longer stimulus values for the older listeners, but the shape of the function is not notably different for younger and older listeners.

Hearing loss effects were seen consistently in the continua that varied the silent interval duration. Hearing-impaired listeners required a longer period of silence to perceive a stop adjacent to a sibilant, as in *Stu* or *ditch*, compared to listeners with normal hearing. This finding may be related to reduced audibility of the sibilance and/or poorer gap detection ability. These results may extend as well to the perception of accented English by listeners with hearing loss. Non-native speakers of English whose first language is Spanish frequently shorten the silent interval preceding /ʃ/ when saying /tʃ/. The implication of the current findings is that listeners with hearing loss may misperceive affricate phonemes in English.

Low-level noise masking was employed in this study in an effort to equate the audibility of speech signals between the young normal-hearing (noise-masked) listeners and the older normal-hearing listeners. Differences in performance between the older normal-hearing listeners and young noise-masked listeners were observed for some of the measures (e.g., crossover points for one VOT continuum in the sentence context) and underscored the presence of age effects that were not confounded by possible differences in high-frequency sensitivity between the younger and older listeners with clinically normal hearing. However, a second group of young normal-hearing listeners without noise masking was also employed to serve as a baseline referent group. Overall, the results indicate that the two young listener groups (with and without noise masking) performed similarly for all conditions. This finding suggests that the use of a low-level noise masker to equate the thresholds of younger and older listeners with normal hearing may not be necessary for investigations of age-related differences in perceptual judgments involving speech continua presented at high suprathreshold levels.

Context effects were observed for most of the speech continua. For several continua, listeners required longer target cues in the sentence context than in isolation, as in *Greg/Creig*, *dish/ditch*, *Sue/Stu*, *Rite/Ride*, and, for older listeners, *buy/pie*. However, for the continua that varied the transition duration (*beat/wheat* and *Bill/Will*), the context effect was attributed to shorter crossover points for words presented in sentence contexts than in isolation. These divergent results are likely due to the nature of the exact sentence context for these stimuli rather than to differential effects of memory. If memory effects played a significant role in these findings, the context effects would be consistent for all stimulus continua, with age-related deficits being exaggerated in all sentence contexts. Rather, the current findings suggest that context effects, when observed, generally show different crossover points for words in the sentence-final position than either isolated words or words in the sentence-initial position, suggesting that effects of specific phonetic sequences play a role in these results.

Overall, the results tentatively suggest that when older listeners with or without hearing loss require longer target temporal cues than younger normal-hearing listeners for words presented in isolation, they are likely to require a similar lengthening of these target cues in sentence contexts. This was observed in five of the eight continua. Moreover, the finding of context effects in seven of the eight continua suggests that patterns of performance observed for the perception of isolated words do not necessarily predict speech perception performance for these same words in the context of a sentence. The current study demonstrates the importance of evaluating the perception of target speech cues in sentence contexts.

ACKNOWLEDGMENTS

This research was supported by a MERIT award from the National Institute on Aging (R37AG09191). The authors

are grateful to Michele Spencer, Jessica Barrett, and Maureen D'Antuono for their assistance with data collection and analysis, and to Ed Smith for implementing the PSIGNIFIT analyses. The authors also appreciate the valuable comments of three anonymous reviewers.

ANSI (**2004**). ANSI S3.6-2004, "American National Standard Specification for Audiometers (Revision of ANSI S3.6-1996), (American National Standards Institute, New York).

Baddely, A. D. (**1986**). *Working Memory* (Oxford University Press, Oxford, England).

Bergeson, T. A., Schneider, B. A., and Hamstra, S. J. (**2001**). "Duration discrimination in younger and older adults," Can. Acoust. **29**, 3–9.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32–48.

Burda, A. N., Scherz, J. A., Hageman, C. F., and Edwards, H. T. (**2003**). "Age and understanding speakers with Spanish or Taiwanese accents," Percept. Mot. Skills **97**, 11–20.

Chen, M. (**1970**). "Vowel length variation as a function of the voicing of the consonant environment," Phonetica **22**, 129–159.

Cruickshanks, K. L., Wiley, T. O., Tweed, T. S., Klein, B. E. K., Klein, R., Mares-Perlman, J. A, and Nondahl, D. M. (**1998**). "Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin: The epidemiology of hearing loss study," Am. J. Epidemiol. **148**, 879–886.

Denes, P. (**1955**). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. **27**, 761–764.

Fisk, A. D., and Rogers, W. (**1991**). "Toward an understanding of age-related memory and visual search effects," J. Exp. Psychol. Gen. **120**, 131–149.

Fitzgibbons, P., and Gordon-Salant, S. (**1994**). "Age effects on measures of auditory duration discrimination," J. Speech Hear. Res. **37**, 662–670.

Fitzgibbons, P. J., and Gordon-Salant, S. (**1995**). "Duration discrimination with simple and complex stimuli: Effects of age and hearing sensitivity," J. Acoust. Soc. Am. **98**, 3140–3145.

Flege, J. E., and Eefting, W. (**1988**). "Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation," J. Acoust. Soc. Am. **83**, 729–740.

Florentine, M., and Buus, S. (**1984**). "Temporal gap detection in sensorineural and simulated hearing impairments," J. Speech Hear. Res. **27**, 449–455.

Fox, R. A., Flege, J. E., and Munro, J. (**1995**). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," J. Acoust. Soc. Am. **97**, 2540–2551.

Gelfand, S., Schwander, T., and Silman, S. (**1990**). "Acoustic reflex thresholds in normal and cochlear-impaired ears: Effect of no-response rates on 90th percentiles in a large sample," J. Speech Hear Disord. **55**, 198–205.

Gordon-Salant, S., and Fitzgibbons, P. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423–431.

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., and Barrett, J. (**2006**). "Age-related differences in identification and discrimination of temporal cues in speech segments," J. Acoust. Soc. Am. **119**, 2455–2466.

Grose, J. H., Eddins, D. A., and Hall, J. W., III (**1989**). "Gap detection as a function of stimulus bandwidth with fixed high frequency cut-off in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1747–1755.

Grose, J. H., Hall, J. W., III, and Buss, E. (**2006**). "Temporal processing deficits in the pre-senescent auditory system," J. Acoust. Soc. Am. **119**, 2305–2315.

Hargus, S. E., and Gordon-Salant, S. (**1995**). "Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment," J. Speech Hear. Res. **38**, 234–243.

He, N.-J., Horwitz, A. R., Dubno, J. R., and Mills, J. (**1999**). "Psychometric functions for gap detection in noise measured from young and aged subjects," J. Acoust. Soc. Am. **106**, 966–978.

Humes, L. E., Lee, J. H., and Coughlin, M. P. (**2006**). "Auditory measures of selective and divided attention in young and older adults using single-talker competition," J. Acoust. Soc. Am. **120**, 2926–2937.

Irwin, R. J., Hinchcliff, L. K., and Kemp, S. (**1981**). "Temporal acuity in normal and hearing-impaired listeners," Audiology **20**, 234–243.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S. (**1956**). "Tempo of frequency change is a cue for distinguishing classes of speech sounds," J. Exp. Psychol. **52**, 127–137.

Lisker, L., and Abramson, A. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Lister, J., Koehnke, J., and Besing, J. (**2000**). "Binaural gap duration discrimination in listeners with impaired hearing and normal hearing," Ear Hear. **21**, 141–150.

MacKay, I. R. A., Flege, J. E., and Piske, T. (**2000**). "Persistent errors in the perception and production of word-initial English stop consonants by native speakers of Italian," J. Acoust. Soc. Am. **107**, 2802.

Munson, B., and Nelson, P. B. (**2005**). "Phonetic identification in quiet and in noise by listeners with cochlear implants," J. Acoust. Soc. Am. **118**, 2607–2617.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Pfeiffer, E. (**1977**). "A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients," J. Am. Geriatr. Soc. **23**, 441–443.

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.

Roup, C. M., Wiley, T. L., Safady, S. H., and Stoppenbach, D. T. (**1998**). "Tympanometric screening norms for adults," Am. J. Audiology **7**, 55–60.

Salthouse, T. A. (**1996**). "The processing-speed theory of adult age differences incognition," Psychol. Rev. **103**, 403–428.

Snell, K. B., and Hu, H.-L. (**1999**). "The effect of temporal placement on gap detectability," J. Acoust. Soc. Am. **106**, 3571–3577.

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.

Tillman, T. W., and Carhart, R. C. (**1966**). "An expanded test for speech discrimination utilizing CNC monosyllabic words: N. U. Auditory Test No. 6," Report No. SAM-TR-66-55, USAF School of Aerospace Medicine, pp. 1–12.

Tun, P. (**1998**). "Fast noisy speech: Age differences in processing rapid speech with background noise," Psychol. Aging **13**, 424–434.

WEDW (waveform display, labeling and editing software), http://www.asel.udel.edu/speech/Spch_proc/software.html (Last viewed January 8, **2008**).

Wichmann, F. A., and Hill, N. J. (**2001a**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293–1313.

Wichman, F. A., and Hill, N. J. (**2001b**). "The psychometric function: II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.

Speech production modifications produced by competing talkers, babble, and stationary noise

Youyi Lu[a] and Martin Cooke
Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, United Kingdom

(Received 2 June 2007; revised 22 July 2008; accepted 27 August 2008)

Noise has an effect on speech production. Stationary noise and babble have been used in the past but the effect of a competing talker, which might be expected to cause different types of disruption, has rarely been investigated. The current study examined the acoustic and phonetic consequences of *N*-talker noise on sentence production for a range of values of *N* from 1 (competing talker) to infinity (speech-shaped noise). The effect of noise on speech production increased with both the number of background talkers (*N*) and noise level, both of which act to increase the energetic masking effect of the noise. In a background of stationary noise, noise-induced speech was always more intelligible than speech produced in quiet, and the gain in intelligibility increased with *N* and noise level, suggesting that talkers modify their productions to ameliorate energetic masking at the ears of the listener. When presented in a competing talker background, speech induced by a competing talker was more intelligible than speech produced in quiet, but the scale of the effect was compatible with the energetic masking effect of the competing talker. No evidence was found of modifications to speech production which exploited the temporal structure of a competing talker.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2990705]

PACS number(s): 43.72.Ar, 43.71.Gv, 43.72.Dv, 43.70.Mn [DOS] Pages: 3261–3275

I. INTRODUCTION

Speech communication frequently takes place in the presence of background noise, conditions known to lead to changes in speech production, collectively known as the Lombard effect (Lombard, 1911). The goal of noise-induced modifications to normal speech production is not yet clear. It has been suggested that by modifying their vocal effort, speakers attempt to maintain a constant level of intelligibility in the face of degradation of the message by the environmental noise source (Summers *et al.*, 1988) and indeed some studies have reported intelligibility gains for "Lombard speech" presented in noise when compared to normal speech in noise (Dreher and O'Neill, 1957; Pittman and Wiley, 2001). However, the issue of how noise-induced speech production changes lead to the intelligibility gain has not yet been addressed.

Many studies have examined the acoustic-phonetic consequences of background noise on speech production (Hanley and Steer, 1949; Dreher and O'Neill, 1957; Charlip and Burk, 1969; Pisoni *et al.*, 1985; Stanton *et al.*, 1988; Summers *et al.*, 1988; Bond *et al.*, 1989; Junqua, 1993, 1996; Letowski *et al.*, 1993; Tartter *et al.*, 1993; Steeneken and Hansen, 1999; Garnier *et al.*, 2006; Pittman and Wiley, 2001; Varadarajan and Hansen, 2006). These studies have converged on a set of primary acoustic changes seen in Lombard speech relative to speech produced in quiet conditions. Specifically, Lombard speech demonstrates an increase in fundamental frequency (F0), speech level, and vowel duration, as well as a flattening of spectral tilt (more energy at higher frequencies). The first and second formant frequencies (F1 and F2) also shift, with the consensus that F1 tends to increase while F2 has been reported to increase (Junqua, 1993) or decrease (Pisoni *et al.*, 1985). In addition, in Lombard speech, energy shifts between different classes of phoneme have been found. Junqua (1993) and Womack and Hansen (1996) reported a shift of energy from consonant to vowel while Hansen (1996) observed energy shifts from semivowel to vowel and consonant. Furthermore, improvements in automatic speech recognition performance under noisy conditions have been reported when Lombard effects have been incorporated into the recognizer (Hansen and Bria, 1990; Hansen, 1994; Chi and Oh, 1996; Hansen, 1996).

Many factors can influence the size of the acoustic changes observed in noise-induced speech modifications. Dreher and O'Neill (1957) reported that increasing the level of the masking noise from 70 to 100 dB resulted in a steady increase in duration from 15% to 31%, and a 6 to 9 dB increase in intensity, over that in quiet. In addition, noise level affects the scale of changes to fundamental frequency, spectral tilt, and formant frequencies (Summers *et al.*, 1988; Tartter *et al.*, 1993). The spectral tilt of the background noise has also been found to influence the Lombard effect. Junqua *et al.* (1998) reported that duration and fundamental frequency tend to increase with noise spectral tilt. Other factors affecting the size of the Lombard effect include the role of the word in a sentence, language spoken, and speaker gender. Patel and Schell (2008) observed larger effects of F0 and duration for information-bearing word types. The effect size was also larger for American English than French (Junqua, 1996). Junqua (1993) reported that the influence of the Lombard effect on vocal effort and F0 was greater for male speakers than for females. In addition, significant inters-

[a]Electronic mail: y.lu@dcs.shef.ac.uk

peaker differences exist in the range of production modifications seen (Stanton *et al.*, 1988; Junqua, 1993).

Acoustic differences between speech produced in quiet and in noise lead to differences in intelligibility. Dreher and O'Neill (1957), Pittman and Wiley (2001), and Summers *et al.* (1988) reported that for the same signal-to-noise ratio (SNR) with isolated words or continuous speech, speech produced in noise is more intelligible than speech produced in quiet. Dreher and O'Neill (1957) suggested that the changes in the spectral and temporal properties of speech which accompany the Lombard effect lead to an improvement in speech intelligibility. Summers *et al.* (1988) also reported that differences in the acoustic-phonetic structure of utterances produced in noise resulted in consistent increases in intelligibility across SNRs and talkers (although only two talkers were used). The magnitude of these effects increased as the environment became more severe. The influence on intelligibility of changes in acoustic parameters such as word length, vocal effort, and consonant-to-vowel energy ratio has also been studied. Howes (1957) showed that intelligibility increases with word length. Pickett (1956) reported that the intelligibility of speech increased with increase in vocal effort when the speech level was low (below 55 dB), remained constant over the range 55–78 dB, but dropped with increasingly forceful shouting (above 78 dB).

One aspect of noise-induced speech production changes which has received little attention is the effect of masking noise with different number of background talkers. Most studies have used stationary noise, though some have employed multitalker babble (Junqua, 1994; Pittman and Wiley, 2001; Garnier *et al.*, 2006). Junqua (1994) discovered that multitalker babble noise led to a larger vowel duration increase as compared to white-Gaussian noise. Garnier *et al.* (2006) demonstrated that increases in voice intensity, spectral energy, and word duration were greater in white noise than in cocktail party noise while mean F0 increased more in cocktail party noise than in white noise. However, wideband noise and multitalker babble did not appear to differentially influence the production of speech (Pittman and Wiley, 2001).

Surprisingly, the effect of an independent single-competing talker on speech production has not been investigated in depth.[1] In this regard, the study of Webster and Klumpp (1962) is relevant. In their study, talker-listener pairs were seated face to face and communicated word lists in conditions of quiet and ambient noise. When there was one background talker-listener pair, the speech level of the foreground talker increased by up to 9 dB, compared to the condition without the background pair. The speaking rate in words per second decreased slightly when the background pair was present. It was also found that the foreground pair made more communication errors when talking at the same time as the competing pair.

In speech perception studies, it is known that a competing talker generates masking effects which differ in two ways from stationary noise. First, at any given global SNR, speech is a far less effective energetic masker than stationary noise (Festen and Plomp, 1990; Simpson and Cooke, 2005). Energetic masking (EM) may be defined as that which results at the auditory periphery due to overlapping excitation patterns. Second, speech on speech produces additional perceptual masking (Carhart *et al.*, 1969) over and above that caused by purely energetic factors, and indeed, this form of "informational masking" (IM) is the dominant effect in determining the intelligibility of a speech target masked by a competing utterance (Brungart, 2001). IM refers to all nonperipheral causes of elevations in masked threshold and includes incorrect assignment of elements in the acoustic mixture to the target speech as well as higher-level factors such as competition for limited attentional resources.

Since speech and noise maskers differ in the degree of EM and IM they produce in speech perception, it is of interest to discover whether they have differing effects on speech production. While the task of speech production in noise differs from speech perception in noise, production might be influenced by perceptual concerns in a number of ways. First, masking noise renders monitoring of a speaker's own productions more difficult, both energetically via loss of information of potential use in feedback and, informationally, due to competing attention. Second, speakers may be able to predict the masking effect of noise in the communicative environment at the ears of their interlocutor. In both cases, alterations to normal speech production might be expected.

The primary purpose of the current study was to determine how noise-induced speech production changes are affected by the degree of EM and IM potential of the noise. To measure the effect of differing amounts of EM and IM, N-speaker babble noises were employed for a range of values of N including $N=1$ (single speaker) and $N=\infty$ (speech-shaped noise). While EM increases with increasing N (Bronkhorst and Plomp, 1992; Simpson and Cooke, 2005), the influence of IM for sentence material is strongest for small N (e.g., $N=2$, Freyman *et al.*, 2004; $N=3$, Carhart *et al.*, 1975) and for $N=8$ for vowel-consonant-vowel tokens (Simpson and Cooke, 2005). Consequently, a number of intermediate values of N were also used in this study, and, in particular, we were interested in the effect of varying N on utterance-level properties such as duration, intensity, and fundamental frequency as well as formant frequencies, energies, and spectral energy distribution at the phonemic level. A further aim was to investigate whether talkers could exploit temporal fluctuations in the noise which are particularly profound for small values of N.

The intelligibility of noise-induced speech is known to increase over speech produced in quiet when noise is added (Dreher and O'Neill, 1957; Summers *et al.*, 1988; Junqua, 1993). A secondary goal of the study was to measure speech intelligibility as a function of the number of talkers and level of background noise. There is still no clear idea of the origin of these intelligibility gains. The current study employed a computational model of EM in an attempt to determine whether the acoustic changes produced by noise-induced speech result from an attempt to reduce the EM effect at the listener's ears.

II. SPEECH PRODUCTION IN NOISE

A. Corpus design

To determine how speech production changes in the presence of widely differing maskers, a corpus of N-talker babble maskers for $N=\{1,2,4,8,16,\infty\}$ was produced. These values were chosen based on an earlier study which measured the masking effect of N-talker babble for a large number of values of N (Simpson and Cooke, 2005). Since one goal of the study was to investigate the role of IM on speech production, talkers produced sentences which were similar in form to those used to produce N-babble maskers. The Grid corpus (Cooke *et al.*, 2006) was used as the source of the masking material, and talkers were asked to read sentences from this corpus. Grid consists of simple six-word sentences such as "lay green with A4 now" or "set white at B8 again." Grid has been used in speech-on-speech tasks and shown to produce large amounts of IM (Cooke *et al.*, 2008), and the noise-intelligibility relation for speech-shaped noise has been measured (Barker and Cooke, 2007). Maskers for the six values of N were presented at 89 dB sound pressure level (SPL), a level in the middle of the range known to induce significant speech production changes (Stanton *et al.*, 1988 used 90 dB; Summers *et al.*, 1988 used 80, 90, and 100 dB; Junqua, 1993 used 85 dB). To examine the effect of noise level for the extreme values of N, the single speaker ($N=1$) and speech-shaped noise ($N=\infty$) maskers were also presented at 82 and 96 dB SPL. Finally, a "quiet" condition was used to provide a reference against which noise-induced speech production modifications could be measured. In summary, talkers produced speech in a total of 11 conditions ($6\times N$ values at 89 dB SPL, $2\times N$ values at 82 dB SPL, $2\times N$ values at 96 dB SPL, and quiet). Symbols used to represent the ten noise conditions here and elsewhere are in the form $N\langle$number of talkers$\rangle$_$\langle$level$\rangle$ so that, for example, $N1$_89 refers to a competing talker background at 89 dB level, while Ninf_96 indicates a speech-shaped noise background at 96 dB level.

B. Sentence lists and maskers

To allow comparison of acoustic and acoustic-phonetic properties, talkers produced the same set of 50 sentences conforming to the Grid syntax in each of the 11 conditions. However, to introduce some variation, each talker produced a different set of 50 sentences. N-babble maskers for the finite values of N were generated by adding utterances drawn at random from the Grid corpus into a 60 s circular buffer until the required babble density was obtained. This approach avoids problems with uneven masking effects which would have occurred if utterances had been added with synchronized start times. One consequence of this strategy was that masking sentences were not synchronized with the talker's productions: background utterances would start at a random point in the sentence and there could be a change in talker during the time allotted to the production of a single utterance. Prior to incorporation into the buffer, leading and trailing silence was removed, and utterances were scaled to have equal root mean square (rms) levels. Masking noise to accompany individual talker productions consisted of 3 s segments of babble drawn at random from the 60 s buffer. Speech-shaped noise was produced by filtering white noise with a filter whose spectrum equalled the long-term spectrum of the Grid corpus, as shown in Fig. 1. Again, a 60 s segment was generated for subsequent random selection.

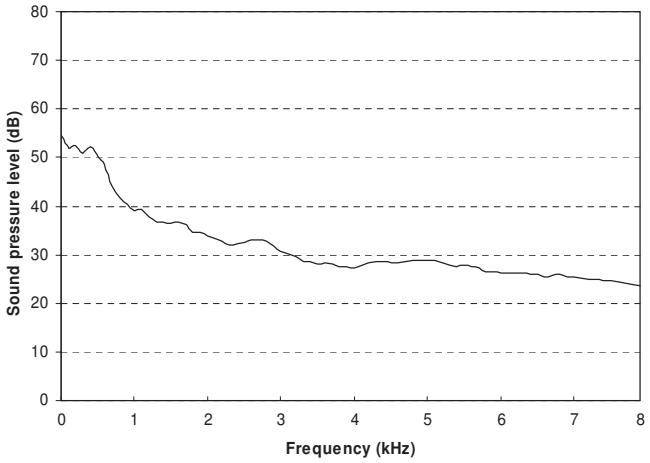


FIG. 1. Long-term average speech spectrum for the Grid corpus.

C. Talkers

Eight native speakers of British English (four males and four females) drawn from staff and students in the Department of Computer Science at the University of Sheffield participated in the corpus collection. All received a hearing test using a calibrated software audiometer which was used to test each ear separately at the six frequencies: 250, 500, 1000, 2000, 4000, and 8000 Hz. One participant had a slight hearing loss (23 dB hearing level (HL)) in one ear at the highest frequency (8 kHz) but was retained for the study. The remaining participants had normal hearing (better than 20 dB hearing level in the range of 250–8000 Hz). Ethics permission was obtained following the University of Sheffield Ethics Procedure. Talkers were paid for their participation.

D. Procedure

Corpus collection sessions took place in an IAC single-walled acoustically isolated booth. Speech material was collected using a Bruel & Kjaer (B & K) type 4190 $\frac{1}{2}$ in. microphone coupled with a preamplifier (B&K type 2669) placed 30 cm in front of the talker. The signal was further processed by a conditioning amplifier (B & K Nexus model 2690) prior to digitization at 25 kHz with a Tucker-Davis Technologies (TDT) System 3 RP2.1. Simultaneously,[2] maskers were presented diotically over Sennheiser HD 250 Linear II headphones using the same TDT system. Speakers wore the headphones throughout, including for the quiet condition. Of course, the use of closed headphones to deliver masking noise can be expected to introduce frequency-dependent own-voice attenuation (Arlinger, 1986; Bořil *et al.*, 2006). Since the current study involves comparison across masking conditions, the constant attenuation characteristics were not considered to be an important factor. However, the closed headphone setup was compared with a com-

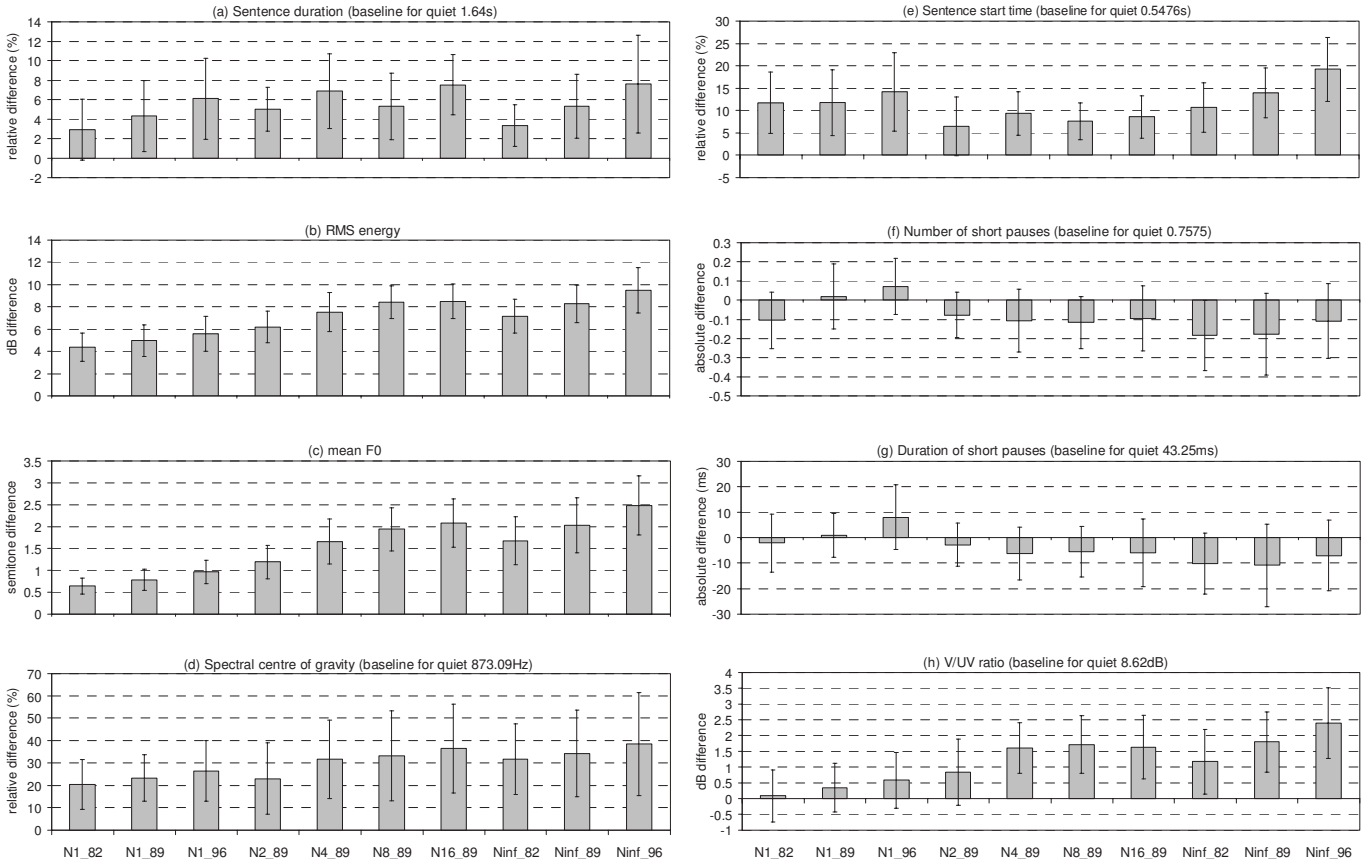


FIG. 2. Differences between acoustic parameter values for each noise condition compared to speech produced in quiet. Where meaningful, "baseline" parameter values in quiet are given in order to provide an absolute reference. Values shown are means over talkers and error bars, here and elsewhere, indicate 95% confidence intervals. Noise conditions are indicated as *N*⟨number of talkers⟩_⟨level⟩.

pensated transmission channel for a subset of conditions. The chief finding was that the recording method was not a significant factor in the speech production modifications measured (see Appendix for details).

Sentence collection and masker presentation was under computer control. Talkers were asked to read out sentences presented on a computer screen and had 3 s to produce each sentence and were allowed to repeat the sentence if they felt it necessary. All the repetitions were saved to allow analysis of the number of "false starts" in the different masking conditions. Prior to saving, signals were scaled to produce a maximum absolute value of unity to make best use of the amplitude quantization range. Scale factors were stored to allow the normalization process to be reversed.

Talkers recorded the 11 conditions over two sessions of 30 min each on two days. They were familiarized with the type of sentences and the task before each collecting session. The three single-talker conditions were combined into a single block as were the three speech-shaped noise conditions, and both sentence order and masker level were randomized within the block. Thus, the 11 conditions were presented in seven blocks, and block order was randomized for each talker.

E. Postprocessing

In order to measure acoustic parameters at the level of individual phonemes, a set of speaker-independent phoneme-level hidden Markov models (HMMs) was built from speech material in the Grid corpus (Cooke *et al.*, 2006) using the HTK HMM toolkit (Young *et al.*, 1999). These models were used to produce phoneme-level transcriptions of the collected utterances via forced alignment using the HVITE tool in HTK. Leading and trailing silent intervals identified via the alignment process were removed. For each talker in each of the 11 conditions, transcriptions of a random selection of 10% of the utterances were manually inspected and found to be accurate.

III. ACOUSTIC AND ACOUSTIC-PHONETIC ANALYSES

A. Utterance-level analysis

Eight acoustic properties were estimated for each utterance. Sentence duration, rms energy, mean fundamental frequency (F0), and spectral center of gravity (CoG) were computed via PRAAT 4.3.24 (Boersma and Weenink, 2005).[3] Sentence start time (i.e., the onset of speech production relative to the onset of the interfering signal) and the number and duration of short pauses (>20 ms) were computed using phoneme-level transcriptions. These latter measures were motivated by the possibility that talkers might avoid overlapping with the background signal, especially in the competing speech conditions. Finally, the voiced-to-unvoiced energy ratio (V/UV ratio) was estimated.

Differences between across-talker means in each background compared to the quiet condition are shown in Fig. 2

TABLE I. Summary of the results of statistical analyses comparing the values of acoustic parameters for speech produced in quiet with speech produced in noise. Column N_89 represents six N-talker conditions ($N=\{1,2,4,8,16,\infty\}$) at a noise level of 89 dB. Columns $N1$ and $N\text{inf}$ represent three level conditions (level=$\{82,89,96$ dB$\}$) for $N=1$ and $N=\infty$, respectively. The final ten columns represent the individual noise conditions as follows. (1–3) $N=1$, levels 82, 89, and 96 dB; (4–7) $N=\{2,4,8,16\}$ at 89 dB; (8–10) $N=\infty$, levels 82, 89, and 96 dB. Symbols "↑" and "↓" represent significant increases or decreases in the parameter in noise over the quiet condition. Significance levels: ***<0.001, **<0.01, and *<0.05.

	Repeated-measures ANOVA			One-sample t-test (test value=0)									
	N_89	$N1$	$N\text{inf}$	1	2	3	4	5	6	7	8	9	10
Sentence duration		* ↑	* ↑			* ↑	** ↑	** ↑	* ↑	** ↑	* ↑	* ↑	* ↑
rms energy	*** ↑	*** ↑	** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑
Mean F0	*** ↑	** ↑	** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	** ↑	*** ↑	*** ↑
CoG	* ↑	* ↑		** ↑	** ↑	** ↑	* ↑	* ↑	* ↑	** ↑	** ↑	* ↑	* ↑
Sentence start time			** ↑	* ↑	* ↑	* ↑		** ↑	** ↑	** ↑	** ↑	** ↑	** ↑
No. of short pauses	* ↑	** ↑											
Duration of short pauses		* ↑											
V/UV ratio	* ↑		* ↑					*** ↑	*** ↑	*** ↑	** ↑	*** ↑	*** ↑

for each of the eight acoustic parameters. The number of talkers and noise level in each background is shown as is the baseline mean for the parameter (that is, the mean value in the quiet condition).

To aid the interpretation of Fig. 2, several statistical analyses were carried out for each acoustic parameter. A repeated-measures analysis of variance (ANOVA) analyzed the effect of the number of talkers (N) in the maskers at the 89 dB level. To determine any interaction effect between N and noise level, a two-way repeated-measures ANOVA with within-subjects factors of N (1, ∞) and masker level (82, 89, 96 dB) was computed. Two further single-factor repeated-measures ANOVAs examined the effects of noise level in the single talker and speech-shaped noise condition. Finally, one-sample t-tests (test value=0) were employed to determine the significance of differences between each masking condition and quiet.

Table I summarizes the results of the statistical analysis for each of the utterance-level acoustic measurements. Many parameters demonstrated significant increases in most of the noise backgrounds compared to quiet (final ten columns of Table I). The most significant effects were for energy (which increased by between 3 and 9 dB relative to quiet) and mean F0 (0.6–2.5 semitones). Spectral CoG increased from the quiet baseline of 870 Hz by 20%–38%. The mean sentence duration in quiet of 1.64 s rose by 2.4%–7.6%, while the pause before speaking increased by 6%–18% from a baseline of 0.55 s in quiet. The V/UV ratio rose in most conditions, from 8.6 dB in quiet by up to 2.4 dB. No significant overall effect of the duration of short pauses or the number of short pauses was found.

For some parameters, the difference between speech produced in quiet and that produced in the presence of noise increased with the number of talkers in the babble (column 2 of Table I). The strongest effect of N was seen for energy and F0, with a lesser effect of spectral CoG and V/UV ratio. However, the effect of N typically reached a plateau at around $N=8$ talkers. For duration, energy, and mean F0, the effect of noise level was similar for the single talker and speech-shaped noise backgrounds (columns 3 and 4 of Table I). Sentence start time (the delay before the talker started speaking following the onset of the noise) increased with level for the speech-shaped noise condition, while the number (and to a lesser extent, duration) of short pauses increased with level in the single-competing talker condition. Similarly, effects of CoG and duration of short pauses were only found in the competing talker conditions, while V/UV ratio increased in the stationary noise conditions. No interaction between the effects of N and noise level was found for any of the parameters.

Figure 2 also indicates the range of talker variation for each parameter and background. While all talkers showed similar changes in energy and F0, significant cross-talker variability is present for the remaining measures.

Increases in duration, mean energy, and F0 in stationary noise were also found in previous studies on Lombard speech using words and short sentences (Dreher and O'Neill, 1957; Pisoni *et al.*, 1985; Summers *et al.*, 1988; Bond *et al.*, 1989; Letowski *et al.*, 1993; Tartter *et al.*, 1993; Steeneken and Hansen, 1999; Garnier *et al.*, 2006). The V/UV ratio increased in most of the N-talker conditions, echoing the findings of Junqua (1993) and Womack and Hansen (1996) for stationary noise. However, a pattern of reduced sentence duration was found by Varadarajan and Hansen (2006), who also reported a decrease in short pause duration, while no such effect was found here. Varadarajan and Hansen (2006) suggested that decreases in sentence and short pause duration could be caused by a sense of urgency on the part of the speaker, which occurred due to the constant exposure to the background noise. Here, the fact that noise was presented only when the talker was due to speak might account for the differences.

B. Phoneme-level analysis

Prior to phoneme-level analysis, all of the postprocessed utterances were normalized to have equal rms energy. Individual phonemes of the utterances were segmented via the phoneme-level transcriptions. Phonemes were grouped into the six categories {vowel, diphthong, liquid, fricative, plo-

TABLE II. Phoneme categories.

Vowel	i:, ɪ, e, u:, æ
Diphthong	eɪ, aɪ, aʊ,
Liquid	w, l, r
Fricative	f, s, v, z, ð
Plosive	p, t, k, b, d, g
Nasal	n

sive, nasal}, as shown in Table II. Phonemes which had fewer than 500 instances were not used. On average, around 3000 instances of each phoneme were employed.

Duration was measured for all phoneme instances, while spectral CoG was measured for all apart from the plosives, whose more complex spectrotemporal development precluded a meaningful measurement. Spectral tilt was computed for all the vowels. It is important to note that due to the limited number of contexts present in the Grid corpus, the phoneme instances used in this analysis should not be regarded as prototypical. For instance, in Grid, the /æ/ vowel can only be found in the word "at," most of which were reduced to schwa in this context. As a consequence, a formant analysis (frequencies, energies, and bandwidths) was undertaken solely for the vowels /i:/, /ɪ/, /e/, and /u:/ in the words "green," "bin," "red," and "soon," respectively. Frequency and energy values were computed as the average of the central three frames in each vowel instance. All of the measurements apart from spectral tilt were computed using the Praat program V 4.3.24 (Boersma and Weenink, 2005). For spectral tilt, the spectrum of an entire phoneme instance was divided into ten energy bands following Stanton *et al.* (1988). Spectral tilt was estimated as the slope of the best linear fit to the ten log energy values. Individual talker and overall measurements were computed for each phoneme. Measurements were obtained by averaging the differences between the phoneme instances of the utterances from each of the ten *N*-talker conditions and the instances in the same position of the same speech sentences from the quiet condition. Individual and overall measurements for all the acoustic properties are expressed as relative percentage differences from quiet, apart from formant frequency and bandwidth, which were expressed as Hertz difference, and energy, which used difference in decibels.

Figures 3 and 4 display the quantitative results of the phoneme-level analysis. To enhance the readability of the plots, results have been averaged across subsets of the ten noise backgrounds. In general, changes in noise backgrounds over quiet were found, and stronger effects were observed for larger number of background talkers and for higher noise levels.

Compared to quiet, increases in *N* and masker level led to an increase in the duration of most sound types apart from the fricative /f/ and the nonalveolar plosives, for which a slight shortening was observed. Increases in spectral CoG were seen for all sounds. For most, the increase was substantially larger than 25%, although the fricatives /f/ and /s/ showed only modest increases. Similar findings for the duration and CoG of vowels have been reported (Junqua, 1993; Stanton *et al.*, 1988; Garnier *et al.*, 2006). Vowel spectral tilt

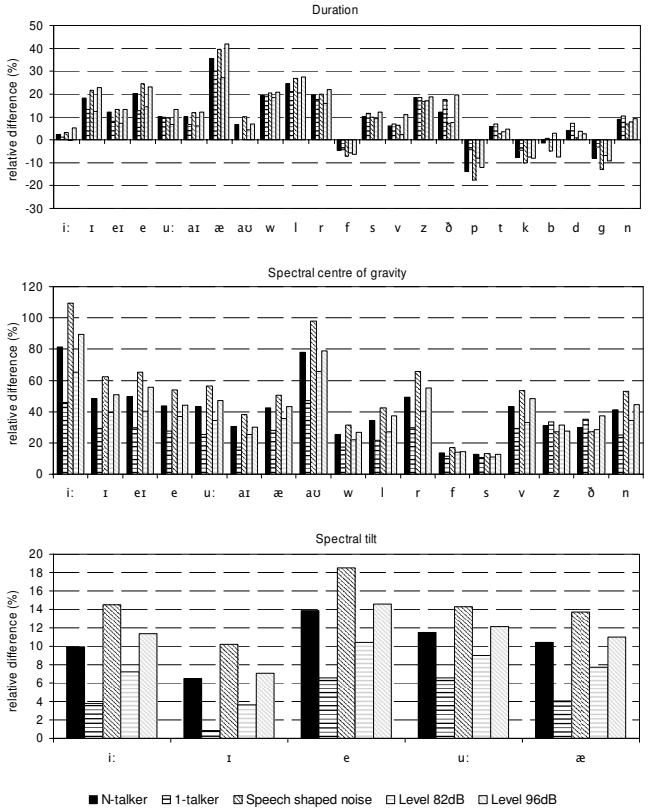


FIG. 3. Phoneme-specific differences in duration (top), spectral CoG (middle), and spectral tilt (bottom) in noise and quiet conditions. For ease of display, the noise conditions are grouped into five subsets. *N*-talker: all ten noise backgrounds; one talker: *N*1_82, *N*1_89, and *N*1_96; speech-shaped noise: *N*inf_82, *N*inf_89, and *N*inf_96; level 82 dB: *N*1_82 and *N*inf_82; level 96 dB: *N*1_96 and *N*inf_96.

became flatter in all conditions, with differences in degree between the vowels. Other studies have reported a similar pattern (Pisoni *et al.*, 1985; Summers *et al.*, 1988; Varadarajan and Hansen, 2006).

In addition, similar statistical analyses to those used for utterance-level parameters in Sec. III A were carried out for formant frequencies, energies, and bandwidths for each vowel. For speech produced in noise, F1 frequency increased significantly by up to 100 Hz. Such effects were stronger for speech-shaped noise, compared to a competing talker background, for all the vowels. F2 and F3 frequencies fell by as much as 60 and 80 Hz, respectively, but these tendencies were only statistically significant for the vowels /i:/ and /ɪ/. For F2 and F3 frequencies, no significant differences were found between competing talker and speech-shaped noise. Increases in vowel F1 frequency were also seen in earlier studies (Pisoni *et al.*, 1985; Summers *et al.*, 1988; Bond *et al.*, 1989; Junqua, 1993; Garnier *et al.*, 2006). For F2, Junqua (1993) reported increases for females while Pisoni *et al.* (1985) found the opposite for both males and females. Other studies (Summers *et al.*, 1988; Bond *et al.*, 1989; Garnier *et al.*, 2006) demonstrated a large amount of vowel and utterance-dependent pattern of F2 frequency change. Junqua (1993) suggested that the F3 frequency of vowels tends to remain constant in noise.

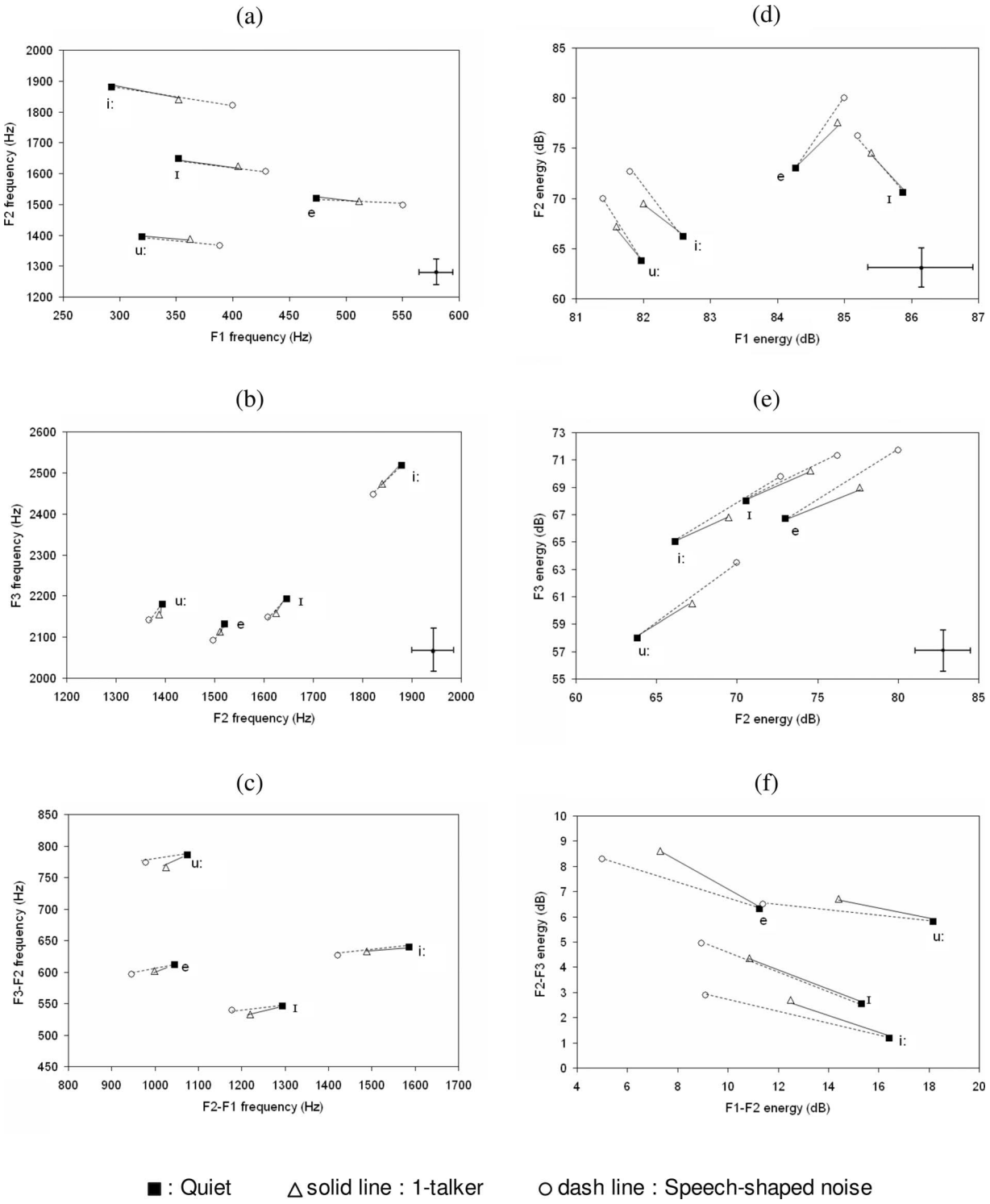


FIG. 4. Formant frequencies (left) and energies (right) for the vowels in "green," "bin," "red," and "soon," for speech produced in quiet and noise. In each case, values are averages taken from the central three frames over all instances of the vowels. For clarity, averages across the three single-talker and three speech-shaped noise conditions are shown. Error bars in lower-right corner indicate 95% confidence intervals.

Significant increases in F2 and F3 energy for the vowels compared to the quiet condition were measured. Such effects were significantly stronger for speech-shaped noise compared to the competing talker background. F1 energy changed little. The overall effect of formant energy changes is consistent with observed changes in spectral tilt.

Furthermore, significant increases in F1 bandwidth and decreases in F2 and F3 bandwidths for all the vowels compared to the quiet condition were found. For most of the vowels, changes in F1 and F2 bandwidths tended to be significantly larger for speech-shaped noise compared to competing talker background while for F3 bandwidth, such tendencies were only significant for the vowels /i:/ and /ɪ/. The changes in F2 and F3 bandwidths for speech produced in noise are consistent with those reported in Hansen and Bria (1990). For F1 bandwidth, Hansen and Bria (1990) found an

increase for /i:/ and /ɪ/ and a decrease for /e/ while Junqua (1993) suggested a decreasing tendency for most of the vowels.

C. Correlation analysis

The above analyses treat speech production changes as independent of each other, but it is possible that correlated changes exist in acoustic parameters such as F0 and F1 frequencies as a result of speech energy changes. Correlations between energy and both F0 and F1 frequencies were investigated. The Pearson correlation coefficient between energy and F0 and energy and F1 frequency was computed independently for all voiced segments. To arrive at a single correlation measure, the weighted mean of segment-based correlations was derived, with weights given by segment duration. For energy versus F0, there was a slight but significant *decrease* ($p<0.05$) in correlation in most of the noise conditions compared to quiet ($r=0.37$). The correlation decreased significantly with an increasing number of background talkers [$F(2.9,20.1)=3.982$, $p<0.05$]. For energy and F1 frequency, significantly *increased* ($p<0.05$) correlation was found in all noise conditions compared to quiet ($r=0.23$). Correlations also increased with the number of talkers [$F(3.8,26.4)=6.504$, $p<0.01$].

D. Discussion

The current results generally confirm the effects of stationary noise on speech production found in previous studies, both at the level of overall acoustic parameter values and for individual phoneme classes. More importantly, they demonstrate for the first time the effect of the number of talkers making up the background babble, including the case of a single talker. For nearly all of the parameters where there is a significant difference between speech produced in stationary noise and in quiet, there is a similar, but smaller, effect when a single talker speaks in the background while speech is produced. Similarly, changes in noise level which have an effect in the stationary noise case tend also to affect the single-talker case. The effect of intermediate background conditions (i.e., multitalker babble for more than one talker) usually falls somewhere between the two extremes. For all parameters, no interaction between the effects of noise level and the number of background talkers was present. One interpretation of these results is that the Lombard effect is influenced by both noise level and number of background talkers, acting independently.

For those parameters which might be expected to reflect the differences in information conveyed by the background, namely, sentence start time and the statistics of short pauses, some small differences were found. There were more pauses longer than 20 ms in the single-talker background than in the other conditions. The pause prior to speaking was longer in the single-talker background than for most of the babble conditions, although the pause was slightly shorter than in the stationary noise case. It is possible that the noncommunicative task limited the scope for such effects.

Some acoustic effects might be the consequence of intentional changes while others may be secondary, caused by articulatory constraints. For example, as pointed out by Gramming *et al.* (1988), the raising of subglottal pressure in order to create a louder voice causes an increase in F0. On the other hand, it is also possible that in the production of high-pitched voice, SPL is raised due to a larger number of speech pressure cycles per time unit resulted from the increase in F0. In addition, the wider jaw opening in order to increase sound amplitude induces an increase in the first formant frequency (Lindblom and Sundberg, 1971). In the current study, correlations between F0 and energy actually decreased in noise, although F1 frequency and energy became more correlated. Thus, it is possible that speakers were using intentional changes in both energy and F0 in response to noise. It is likely that other factors such as physiological and semantic constraints on possible F0/F1 values and range also limit the extent to which speakers can manipulate these parameters independently.

IV. INTELLIGIBILITY OF NOISE-INDUCED SPEECH

A. Motivation

Speech produced in the presence of noise can lead to increases in intelligibility over speech produced in quiet mixed with equivalent noise tokens at the same SNR (Dreher and O'Neill, 1957; Summers *et al.*, 1988; Junqua, 1993). The speech material collected in the current study employed a wider range of noise backgrounds, allowing several new issues to be explored. First, the general finding that the effect of noise on certain acoustic parameters tended to increase with both noise level and number of talkers (N) suggests that any intelligibility gains may also be influenced by noise level and N. Experiment I measured speech intelligibility as a function of noise level and N for noise-induced speech compared to speech produced in quiet with added noise.

When faced with the task of communicating in the presence of a single-competing talker, talkers might adopt strategies to reduce both the EM and IM components at the ear of the listener. Two further experiments explored these possibilities. In experiment II, listeners were presented with utterances masked by a competing talker. The intelligibility of utterances produced in quiet was measured and compared to that of the same set of utterances induced by a competing talker when presented in the background of the inducing competing talker maskers. Any intelligibility gains in the latter "matched" case might be interpreted as resulting from a talker's awareness of the IM effect of the competing utterance. However, increases in intelligibility could also be derived from reductions in EM due to acoustic changes in the competing speaker-induced utterances. Experiment III attempted to distinguish the two hypotheses by comparing the intelligibility of speech produced in the presence of a competing talker when presented in the matched competing talker background with the same utterances presented in an unmatched competing talker background. If talkers are sensitive to the IM potential of a specific competing utterance rather than the EM properties of speech in general, listeners should produce higher scores in the matched condition.

B. Experiment I: Sentences in stationary noise

1. Listeners

Twelve native speakers of British English (nine males and three females) drawn from the undergraduate and postgraduate population of the Department of Computer Science at the University of Sheffield took part in experiment I. All subjects received a hearing test using the same software and procedure, as described in Sec. II C. All had normal hearing apart from one participant with a hearing level of 25 dB in one ear at 8 kHz. This subject was retained for the study. Ethics permission was obtained following the University of Sheffield Ethics Procedure.

2. Stimuli

Utterances collected in quiet and in the presence of noise were presented in a background of stationary speech-shaped noise. Five sets of 100 utterances were used, corresponding to speech produced in quiet, in a background of a competing talker at levels of 82 and 96 dB SPL, and in a background of stationary noise at 82 and 96 dB SPL. In all five conditions of experiment I, utterances were mixed with a speech-shaped noise masker at an overall SNR of −9 dB, a value chosen on the basis of pilot tests to reduce ceiling and floor effects. Prior to mixing, target utterances were scaled to have the same rms level. Maskers were gated on and off with the endpointed utterances, and the mixed signals were scaled to a level of approximately 68 dB SPL.

3. Procedure

Experiment I took place in an IAC single-walled acoustically isolated booth. Stimuli presentation and results collection was controlled by a computer program. Stimuli were presented diotically over Sennheiser HD 250 Linear II headphones via a TDT System 3 RP2.1. Listeners were given instructions to identify in each noisy utterance the letter and digit keywords. This they did via a computer keyboard whose keys were selectively activated to minimize keying errors. For consistency with later experiments, in which the color keyword was used to identify the target utterance, sentences within each condition were organized into four blocks by color keyword. Condition order was balanced across listeners while both color blocks and utterance order within blocks were randomized for each listener. The experiment took place in a single session which was preceded by a short practice. In addition, four practice tokens were added to the start of each condition. Listeners were unaware of these tokens and they were not scored. The entire session required around 30 min to complete.

4. Results

For utterances produced in quiet and presented in speech-shaped noise, listeners obtained a mean keyword identification score of 42%. However, for the four conditions involving the identification of utterances produced in a noise background, keyword scores were substantially higher. As shown in Fig. 5, the increase in scores for noise-induced speech ranged from 9 to 25 percentage points. These increases were statistically significant ($p<0.01$ in the single-talker 82 dB condition; $p<0.001$ in the other three conditions). The two single-talker backgrounds led to the smallest improvements, and in both the single-talker and stationary noise backgrounds, the gain in intelligibility increased with noise level. Among the four noise-induced speech conditions, a two-way repeated-measures ANOVA with factors of $N=\{1,\infty\}$ and level$=\{82,96$ dB$\}$ found a significant effect of N [$F(1,11)=27.276$, $p<0.001$] and noise level [$F(1,11)=8.278$, $p<0.05$]. The N by noise level interaction was not significant ($p>0.2$).

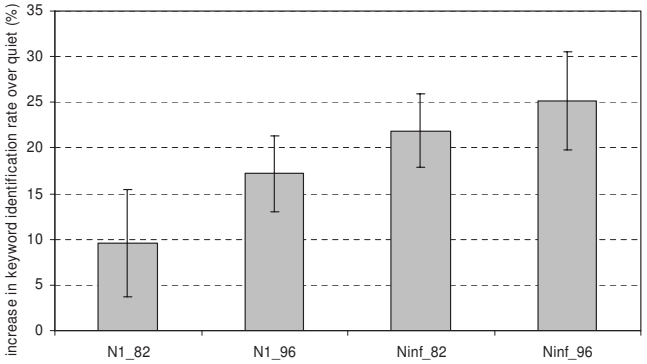


FIG. 5. Keyword identification rates for noise-induced speech over speech produced in quiet when added to speech-shaped noise (experiment I). The baseline keyword identification score for utterances produced in quiet is 42%.

C. Experiment II: Sentences in competing utterances

1. Listeners, stimuli, and presentation

Listeners who took part in experiment I also took part in this experiment. Four conditions tested the identification of keywords in utterances when presented in a competing speaker background. In two conditions, listeners heard speech produced in quiet conditions added to other speech material produced in quiet, drawn from the same corpus (Cooke *et al.*, 2006). In the other two conditions, listeners heard speech that was produced in a competing speech background added to that competing speech background. These "speech-induced" conditions were drawn from those collected as described in Sec. II and corresponded to the 82 and 96 dB background levels. Both "quiet" conditions were identical apart from the choice of sentences used for the background. Two conditions were used to enable the same set of speech maskers to be used in the speech-induced and quiet conditions.

As for experiment I, 100 utterances were used for each condition. For this experiment, sentences contained no keywords in common with those of the masker. Sentences were added so that the target to masker ratio was −9 dB, a value chosen on the basis of pilot experiments, and maskers were gated on and off with the endpointed target sentence. Due to the approach taken to the generation of competing speech maskers as described in Sec. II, the start of a sentence did not necessarily coincide with the start of a sentence in the masker. In this respect, the two-talker scheme was different from those used in IM experiments (e.g., Brungart, 2001; Cooke *et al.*, 2008).

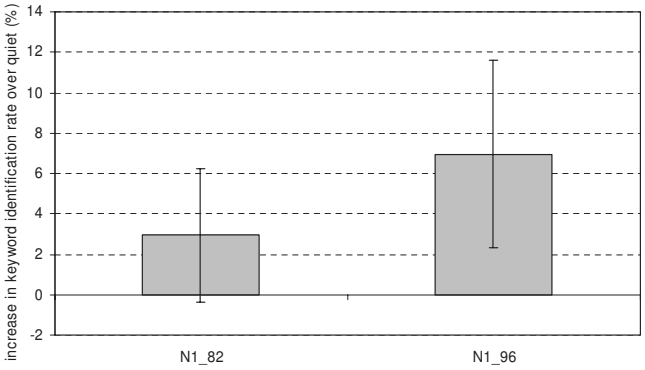


FIG. 6. Keyword identification rates for utterances induced by a speech background over utterances produced in quiet when added to the inducing speech (experiment II). The baseline keyword identification scores for utterances produced in quiet are 81% and 78%, respectively.

The stimulus presentation setup was as described in experiment I. Since this task involved identifying a target in a very similar masker, listeners required information to distinguish the target and masker sentences. The color keyword was used to indicate which utterances listeners had to attend to. The corpus contains four color keywords, so stimuli were organized into four blocks within each condition. At the start of each block, listeners were instructed (via the computer screen) to identify the letter and digit in the sentence containing a given color.

2. Results

Figure 6 displays the difference in keyword identification rates between the speech-induced and quiet utterances for the two levels 82 and 96 dB. While the speech-induced utterances are more intelligible for both levels, only the 96 dB case reaches statistical significance at the 0.05 level [$t(11)=1.756$], suggesting that speech produced in sufficiently intense backgrounds containing a single-competing talker is more intelligible than speech produced in quiet when added to the same competing talker material. This finding extends that of experiment I to a highly nonstationary masker. However, the absence of an effect for speech produced in less intense backgrounds calls into question the extent to which this effect is due to an attempt by the speaker to minimize the degree of IM at the ear of the listener.

D. Experiment III: Induced speech in matched and unmatched backgrounds

1. Listeners, stimuli, and presentation

Listeners who participated in experiments I and II also took part in this experiment. Experiment III compared two conditions, one in which the target material consisted of speech induced by other speech was presented in the background of the inducing speech material ("matched") and one in which the same target speech was presented in "unmatched" backgrounds. Target speech consisted of utterances collected as described in Sec. II in the presence of a competing talker presented at 89 dB SPL. All other stimulus construction and presentation details were the same as for experiment II (100 utterances, −9 dB target-to-masker ratio, and presentation of targets blocked by color keyword). Experiments II and III were performed in sequence in the same session, which lasted approximately 30 min.

2. Results

Keyword identification score in the matched condition was 1.4 percentage points higher than in the unmatched condition (score of 86%). However, this failed to reach statistical significance at the 95% level ($p=0.08$). This outcome suggests that, in this task, talkers do not modify their productions in response to the details of a specific competing utterance.

E. Discussion

The three perceptual experiments here explored the extent to which the presence of competing speech and stationary noise influences the intelligibility of speech productions. Experiment I confirmed previous findings on the increased intelligibility of speech produced in stationary noise backgrounds (Dreher and O'Neill, 1957; Summers *et al.*, 1988) and extended these results to single-talker maskers. The size of intelligibility gains was closely correlated with the extent of acoustic changes measured in Sec. III: stationary noise backgrounds and intense background level both resulted in larger intelligibility gains than single-talker backgrounds and less intense backgrounds. However, all backgrounds tested resulted in significant gains in intelligibility.

Experiments II and III employed maskers designed to invoke large amounts of IM to explore the possibility that talkers modify their production strategy dynamically in response to the presence of competing speech. Experiment II demonstrated that speech produced in an intense competing speech background was more intelligible than speech produced in quiet when presented in the same background. However, for speech produced in a less intense background, no such difference was found, suggesting that EM rather than IM is dominant since the lower background intensity during production (82 dB SPL) is still relatively strong and could be expected to produce IM effects. It seems likely that similar principles as those leading to modifications in production for speech produced in stationary noise backgrounds operate in the competing speech condition.

The results of experiment III do not support the idea that talkers modify their productions in response to the details of individual competing utterances in order to improve intelligibility at the ear of the listener. There was no significant difference in identification scores between speech produced in the speech backgrounds for the maskers which induced the utterances compared to the same induced utterances presented with random speech maskers.

V. DOES NOISE-INDUCED SPEECH OFFER MORE GLIMPSING OPPORTUNITIES?

A. Motivation

While the finding that noise-induced speech is often more intelligible when presented in noise has been reported in studies dating back many years (Dreher and O'Neill,

1957) and has been confirmed here, little effort has been directed toward an explanation of the intelligibility gain. Here, we test the hypothesis that the intelligibility of noise-induced speech is related to the availability of "glimpses" of speech at the ear of the listener in the presence of noise. Glimpses of a signal are defined as those connected regions in its spectrotemporal representation over a certain minimum "area" calculated from the number of spectrotemporal "pixels" and where each spectrotemporal pixel has a local SNR larger than a threshold (Cooke, 2006). This hypothesis is grounded in the EM produced by the masker and differs from a pure EM explanation in that glimpses incorporate the idea that listeners only have access to spectrotemporal regions which are sufficiently dominant in both local SNR and spectrotemporal extent to allow them to stand out above the masker (Cooke, 2006).

B. Glimpse measures

Two glimpsing statistics were measured for the signal mixtures used in the intelligibility experiments described in the previous section. One, "glimpse area," is the number of spectrotemporal points where the glimpse criteria described above hold. Since glimpse area will typically increase with signal duration, "glimpse proportion" was also computed, defined as the proportion of spectrotemporal points which meets the glimpse criteria. This latter measure is independent of duration and helps to distinguish simple speech production processes which improve glimpsing opportunities by slowing speech rate from those which reallocate energy in time and frequency to improve glimpsing opportunities.

Computation of glimpse measures was based on a spectrotemporal excitation pattern (STEP) representation formed for the target and masker independently. A STEP is produced by first passing the time-domain signal through a 64 channel gammatone filterbank, smoothing the Hilbert envelopes, integrating the energy into 10 ms frames, followed by log compression. More details of the computation can be found in Cooke (2006). Following Cooke (2006), a minimum area of 5 and a local SNR of −5 dB were used here.

C. Results

Figure 7 shows the two glimpse measures for each of the conditions used in the experiments of the previous section. For the stationary noise conditions corresponding to experiment I, significantly more glimpses (as measured by both area and proportion) were produced by the noise-induced speech than for speech produced in quiet ($p<0.01$). Stationary noise maskers produced more glimpses than the competing speech [$F(1,7)=14.501$, $p<0.001$ for area; $F(1,7)=10.513$, $p<0.01$ for proportion] while the effect of an increase in noise level was significant only for the glimpse area measure [$F(1,7)=6.438$, $p<0.05$]. Regarding the two competing talker conditions of experiment II, both show significantly more glimpses than speech produced in quiet when measured in terms of glimpse area [$t(7)=3.398$, $p<0.05$ for the less intense condition; $t(7)=3.780$, $p<0.01$ for the more intense condition] while there is a small increase in glimpse

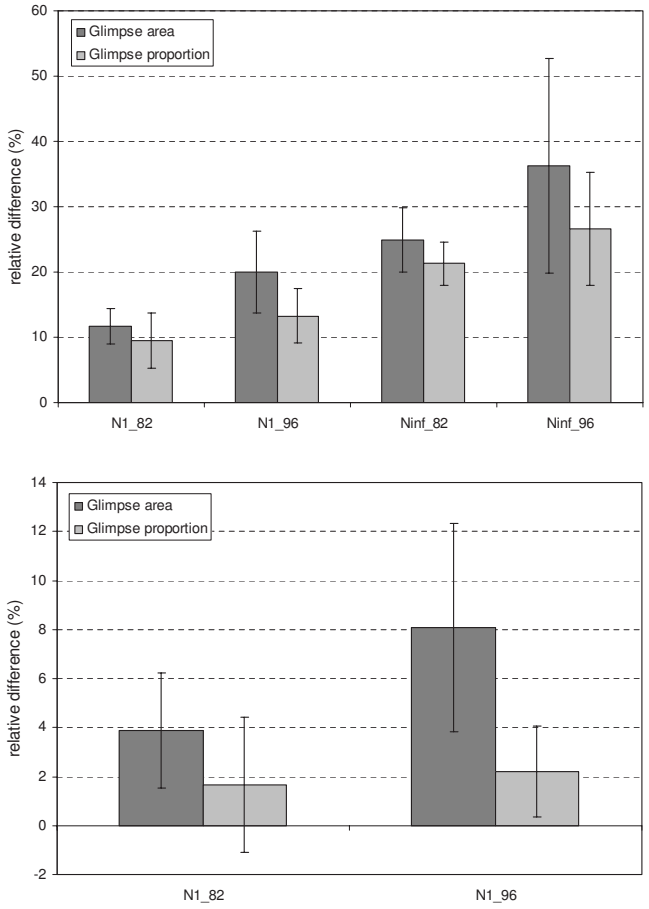


FIG. 7. Glimpse area and proportion for the listening conditions of experiments I (top) and II (bottom) expressed as percentage increase in area or proportion over speech produced in quiet.

proportion for the more intense condition [$t(7)=2.384$, $p<0.05$]. Finally, as was the case for intelligibility, no significant effect was found for experiment III.

Overall, the results are strikingly similar to those for intelligibility, as illustrated by Fig. 8 which plots relative intelligibility gains for listeners against relative increases in the two glimpse measures. Both measures are highly correlated with listener intelligibility gains, suggesting that noise-

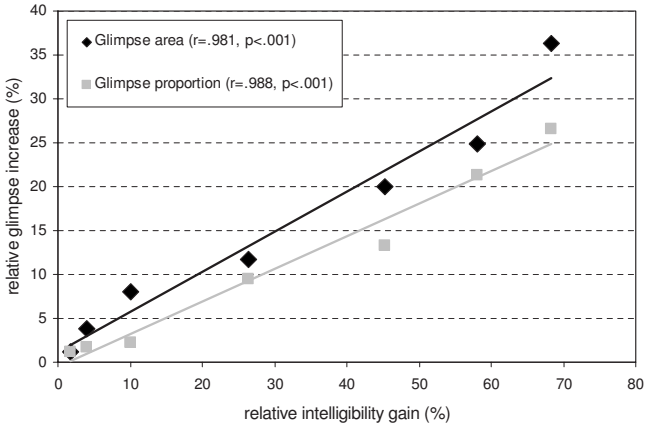


FIG. 8. Relation between increases in the two glimpse measures and increase in intelligibility for experiments I–III together with least-squares fits.

induced speech is more intelligible than speech produced in quiet because the articulatory manipulations lead to a release from EM.

Of the two glimpse measures, significantly larger increases in glimpse area over glimpse proportion are found [$F(1,7)=4.102$, $p<0.05$; $F(1,7)=7.397$, $p<0.05$] for the conditions of both experiments I and II. This is presumably due to the tendency of noise-induced sentences to increase in duration. No significant correlation was found between utterancewise measures of duration and glimpse proportion in any of the noise-induced conditions, suggesting that speakers use both a slower speaking rate, to increase the overall number of glimpses, and other (mainly spectral) modifications in order to increase the proportion of glimpses available for the hearer. However, in explaining listener performance, there is no clear basis to prefer glimpse area over glimpse proportion.

VI. GENERAL DISCUSSION

A. Noise-induced speech and energetic masking

A number of reliable and consistent acoustic modifications occur when speech is produced in the presence of noise. The main effects—increases in F0, energy, and spectral CoG—confirm those found in previous work using multitalker babble and stationary noise. The study extends the scope of noise-induced production effects to single-talker interfering speech, which was also found to be capable of producing significant acoustic changes compared to speech produced in a quiet background. Increases in F0, energy, and CoG grew as the number of talkers in the background increased, asymptoting at around 8–16 talkers. These results demonstrate that the extent of acoustic modifications is largely correlated with both the intensity of the background signal and number of background talkers. This suggests that noise-induced speech production changes are dependent on the overall EM capacity of the background signal since EM is a function of both overall noise level and number of background talkers: a competing talker is a far less effective energetic masker than a broadband noise when both are presented at the same SNR (Festen and Plomp, 1990), and EM increases with the number of background talkers (Bronkhorst and Plomp, 1992; Simpson and Cooke, 2005).

Experiment I demonstrated that noise-induced speech was more intelligible when presented in stationary noise than speech produced in quiet, extending previous findings for stationary and multitalker babble backgrounds. Interestingly, those backgrounds which resulted in the largest acoustic modifications also produced the biggest increases in intelligibility, suggesting that speakers modify their productions in response to the adversity of the background. Indeed, the result of production modifications is to increase the number and proportion of opportunities to glimpse the target speech in noise, and the increase in such opportunities is very closely correlated with listener keyword identification performance. Thus, the potential for EM leads to articulatory modifications whose acoustic consequence is to cause a release from masking, and the more masking potential that exists, the greater the eventual release. These findings support Lindblom's suggestion that speakers compensate for environmental conditions (Lindblom, 1990).

B. Basis for the increased intelligibility of noise-induced speech

It is not clear how the acoustic consequences of changes in speech production lead to increased intelligibility. While it is evident that the overall increase in intensity of noise-induced speech produces a release from EM, this cannot account for the intelligibility gains observed here since all utterances were normalized to have the same SNR when presented alongside maskers. However, speakers can employ a number of other strategies to improve the SNR at the ear of the listener. For instance, a decrease in speaking rate provides more opportunities to glimpse acoustic information useful for phonetic distinctions. The largest increase in utterances duration of around 7% in the most adverse backgrounds might have contributed to the overall improvement in intelligibility, but it is unlikely to be responsible for the entire increase since the results of the glimpsing analysis showed that the *proportion* of the spectrum lying above the masker also increased for the noise-induced conditions.

Many of the acoustic consequences of noise-induced speech are compatible with an overall shift in the energy balance from lower to higher frequencies. For the vowels, increases in fundamental frequency, spectral CoG, and energy for the second and third formants are reflected in a flattening of spectral tilt. One consequence of this shift to higher frequencies is a certain degree of masking release in the presence of the maskers employed in this study, whose mean spectrum was speech shaped. However, vowel formant frequencies became more "central," with increases in F1 and decreases in F3. Of course, there are articulatory limits to the range of speech production modifications possible, and some of the acoustic changes observed may be epiphenomena associated with other manipulations such as increased effort and vocal stress.

The issue of whether speakers actively attempt to place spectral information in locations where it is less likely to be masked merits further study. Formant frequency changes are also found when talkers are asked to speak clearly (Chen, 1980; Picheny *et al.*, 1986; Krause and Braida, 2004; Smiljanic and Bradlow, 2005). These studies categorized vowels as tense or lax, corresponding to /i:, u:/ and /ɪ, e/ here. No consistent trends were found for the first three formant frequencies of tense vowels, although Chen (1980) reported that tense vowels clustered more tightly in vowel formant space in clear than in conversational speech. Picheny *et al.* (1986) and Krause and Braida (2004) reported increases in F1 and F2 of the lax vowel /ɪ/ of up to 50 and 200 Hz, respectively.

C. Speech changes produced by a competing talker

One of the motivations for the current study was to determine how the presence of a competing talker affects speech production. One possibility is that competing speech material might disrupt the speech production process of the

talker, resulting in false starts, hesitations, and other dysfluencies. The speech material used in this study was deliberately chosen to be similar to that introduced in the background in order to provoke such effects. Some disrupting influence of the competing talker background was found: the number and duration of short pauses increased with intensity while no similar effects were seen for the stationary noise backgrounds. Furthermore, the number of false starts was larger in the intense single-talker background than in quiet [$t(7)=2.646$, $p<0.05$]. However, these effects were small and the overall number of short pauses was not significantly greater than in a quiet background.

A second potential influence of competing speech is on the talker-listener communication process: the talker might anticipate the IM effect of two similar utterances at the ear of the listener and employ strategies to reduce the extent of IM. Experiment II demonstrated that utterances produced in the presence of an intense competing talker were more intelligible than utterances produced in quiet conditions when presented in speech backgrounds. For speech produced with a less intense talker, there was no significant gain over quiet. These findings suggest that it is primarily EM rather than IM that leads to increased intelligibility since if the latter were at work, some effect in the less intense background would be expected since the production and background levels are closer and lead to more IM for the listeners (Brungart, 2001). Furthermore, no evidence was found of speaking strategies which exploited the temporal fluctuations of specific competing utterances: there was no difference in the intelligibility of speech in the presence of the material which induced it when compared to speech in the presence of other speech material (experiment III). Talkers may be unable to attend to and track competing speech material sufficiently rapidly to modify their own productions in response.

D. Task dependence

While few effects of a competing talker above and beyond EM were found here, it is possible that other tasks might elicit more extensive speech production changes. The task employed in the current study was devoid of communicative intent, and it was possible for speakers to read the prompts on the screen with little regard for intelligibility. Summers *et al.* (1988) found that with no communicative element there was little incentive for speakers to consciously change their speech even with masking noise present in the headphones. Lane and Tranel (1971) indicated that the speaker does not change his voice level to communicate better with himself, but rather with others. Junqua *et al.* (1999) also concluded that the communication factor has a strong influence. Further studies using two-way interactive task with single-talker maskers need to be conducted before ruling out the possibility of both positive effects of active strategies which are sensitive to the local masking conditions and negative effects of attentional deployment to processing an informative background source while speaking.

It is known that the greatest IM effects are found when the target signal and masker are similar (Brungart, 2001; Neff, 1995; Oh and Lutfi, 2000; Kidd *et al.*, 2002). Indeed, although the masking utterances were similar in form to those produced by the talkers, start times were not synchronized, so the chances of similar words overlapping was reduced. It is possible that tasks designed to produce large amounts of IM would give rise to more significant changes in speech production than those observed in the current study.

VII. CONCLUSIONS

Speakers modified their productions in *N*-talker noise backgrounds across a wide range of values for *N*. This they achieved not only by increases in output level but by changes to the fundamental frequency and formant energies which result in an overall increase in spectral CoG. The scale of acoustic modifications increased both with *N* and the level of the background noise, conditions which also result in increases in the EM effect of the noise. Noise-induced speech was more intelligible when presented in stationary noise than speech produced in quiet, and the intelligibility gain increased with *N* and noise level. These findings, coupled with a computer model of EM, suggest that speakers attempt to compensate for the EM effect of the noise on their own speech. In contrast, no IM effects of a competing talker were found, perhaps because the task lacked a communicative element.

APPENDIX: COMPENSATION FOR OWN-VOICE ATTENUATION

To determine whether own-voice attenuation caused by closed headphones was a factor in the current study, a compensation method was introduced. First, the spectral difference of a white noise signal with and without Sennheiser HD 250 Linear II headphones was measured using a B & K type 4100 head and torso simulator equipped with B & K type 4190 $\frac{1}{2}$ in. microphones. An order-32 IIR filter was designed to have a transfer function which was the inverse of the attenuation characteristic produced by the headphones. This filter was implemented on a TDT RP 2.1 processor and compensated for the headphone attenuation in real time.

In order to discover whether the original and the compensated recording method produced similar effects on speech production, a small corpus was collected using the two methods and analyzed at utterance and phoneme level. Eight native speakers of British English (four males and four females) drawn from staff and students in the Department of Computer Science at the University of Sheffield participated in the corpus collection. Eight recording conditions were employed which included quiet, competing talker, eight-talker babble, and speech-shaped noise. Talkers produced the same set of 25 sentences in each of the eight conditions. Maskers for noise conditions were produced as described in Sec. II B and presented at 89 dB SPL. Condition order was randomized for each talker. For the collected utterances, leading and trailing silent intervals identified via the alignment process described in Sec. II E were removed.

Four acoustic properties were estimated for each utterance in each of the eight conditions. Sentence duration, rms energy, mean fundamental frequency (F0), and spectral CoG

were computed as described in Sec. III A. A two-way repeated-measures ANOVA (two recording methods × three noise conditions) was computed for each acoustic parameter. *Post hoc* analysis showed that for all parameters and noise conditions, there was no significant effect of recording method [$F(1,7)=0.893$, $p=0.376$ for duration; $F(1,7)=1.240$, $p=0.278$ for energy; $F(1,7)=0.923$, $p=0.369$ for F0; $F(1,7)=1.055$, $p=0.339$ for CoG]. For the quiet and competing talker conditions, short pauses within each utterance were manually identified and their number and duration computed. Again, the difference in recording setups led to no statistically significant differences [$F(1,7)=0.007$, $p=0.936$ for the number of short pauses; $F(1,7)=0.056$, $p=0.820$ for duration of short pauses].

[1]However, speech produced in the presence of other speech material has been studied in the limited sense of altered auditory feedback (Lee, 1950; Natke and Kalverman, 2001; Stuart *et al.*, 2002; Xu *et al.*, 2004).

[2]Processing delays in the TDT System 3 processor mean that the noise output was slightly delayed (maximum 6 ms) with respect to speech input.

[3]Mean energy was computed via "get intensity decibels." F0 estimates were provided at 10 ms intervals using an autocorrelation-based method (Boersma, 1993) implemented in the PRAAT program. Mean F0 was obtained by averaging all the valid F0 estimates and expressed in semitones. Spectral CoG was computed on the spectrum of an entire utterance via the PRAAT command "get CoG" and expressed in hertz, using a linear frequency axis and power magnitude spectrum.

Arlinger, S. D. (**1986**). "Sound attenuation of TDH-39 earphones in a diffuse field of narrow-band noise," J. Acoust. Soc. Am. **79**, 189–191.

Barker, J., and Cooke, M. P. (**2007**). "Modelling speaker intelligibility in noise," Speech Commun. **49**, 402–417.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proc. Inst. Phonetic Sci. **17**, 97–110.

Boersma, P., and Weenink, D. (**2005**). "Praat: doing phonetics by computer (version 4.3.14) (computer program)," (Last viewed May, 2005) from http://www.praat.org.

Bond, Z., Moore, T., and Gable, B. (**1989**). "Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask," J. Acoust. Soc. Am. **85**, 907–912.

Bořil, H., Bořil, T., and Pollák, P. (**2006**). "Methodology of Lombard speech database acquisition: Experiences with CLSD," LREC 2006 fifth Conference on Language Resources and Evaluation, pp. 1644–1647.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Carhart, R., Johnson, C., and Goodman, J. (**1975**). "Perceptual masking of spondees by combinations of talkers," J. Acoust. Soc. Am. **58**, S35.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Charlip, W. S., and Burk, K. W. (**1969**). "Effects of noise on selected speech parameters," J. Commun. Dis. **2**, 212–219.

Chen, F. R. (**1980**). "Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level," MS thesis, Massachusetts Institute of Technology, Cambridge.

Chi, S. M., and Oh, Y. H. (**1996**). "Lombard effect compensation and noise suppression for noisy Lombard speech recognition," International Conference on Spoken Language Processing, pp. 2013–2016.

Cooke, M. P. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Cooke, M. P., Barker, J., Cunningham, S., and Shao, X. (**2006**). "An audio-visual corpus for speech perception and automatic speech recognition," J. Acoust. Soc. Am. **120**, 2421–2424.

Cooke, M. P., Garcia Lecumberri, M. L., and Barker, J. P. (**2008**). "The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception," J. Acoust. Soc. Am. **123**, 414–427.

Dreher, J. J., and O'Neill, J. (**1957**). "Effects of ambient noise on speaker intelligibility for words and phrases," J. Acoust. Soc. Am. **29**, 1320–1323.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

Garnier, M., Bailly, L., Dohen, M., Welby, P., and Loevenbruck, H. (**2006**). "An acoustic and articulatory study of Lombard speech: Global effects on the utterance," International Conference on Spoken Language Processing, pp. 2246–2249.

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., and Perkins, W. (**1988**). "Relationship between changes in voice pitch and loudness," J. Voice **2**, 118–126.

Hanley, T. D., and Steer, M. D. (**1949**). "Effect of level of distracting noise upon speaking rate, duration, and intensity," J. Speech Hear Disord. **14**, 363–368.

Hansen, J. H. L. (**1994**). "Morphological constrained feature enhancement with adaptive cepstral compensation (MCE-ACC) for speech recognition in noise and Lombard effect," IEEE Trans. Speech Audio Process. **2**, 598–614.

Hansen, J. H. L. (**1996**). "Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition," Speech Commun. special issue on speech under stress, **20**, 151–170.

Hansen, J. H. L., and Bria, O. N. (**1990**). "Lombard effect compensation for robust automatic speech recognition in noise," International Conference on Spoken Language Processing, pp. 1125–1128.

Howes, D. (**1957**). "On the relation between the intelligibility and frequency of occurrence of English words," J. Acoust. Soc. Am. **29**, 296–305.

Junqua, J. C. (**1993**). "The Lombard reflex and its role on human listeners and automatic speech recognizers," J. Acoust. Soc. Am. **93**, 510–524.

Junqua, J. C. (**1994**). "A duration study of speech vowels produced in noise," International Conference on Spoken Language Processing, pp. 419–422.

Junqua, J. C. (**1996**). "The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex," Speech Commun. **20**, 13–22.

Junqua, J. C., Fincke, S., and Field, K. (**1998**). "Influence of the speaking style and the noise spectral tilt on the Lombard reflex and automatic speech recognition," International Conference on Spoken Language Processing, pp. 467–470.

Junqua, J. C., Fincke, S., and Field, K. (**1999**). "The Lombard effect: A reflex to better communicate with others in noise," Acoustic, Speech, and Signal Processing, 1999 (ICASSP'99) Proceedings, Vol. **4**, pp. 2083–2086.

Kidd, G., 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.

Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**, 362–378.

Lane, H. L., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Lee, B. S. (**1950**). "Effects of delayed speech feedback," J. Acoust. Soc. Am. **22**, 824–826.

Letowski, T., Frank, T., and Caravella, J. (**1993**). "Acoustical properties of speech produced in noise presented through supra-aural earphones," Ear Hear. **14**, 332–338.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, The Netherlands), pp. 403–439.

Lindblom, B., and Sundberg, J. (**1971**). "Acoustical consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Am. **50**, 1166–1179.

Lombard, E. (**1911**). "Le Signe de l'Elevation de la Voix (The sign of the rise in the voice)," Ann. Maladiers Oreille, Larynx, Nez, Pharynx (Annals of diseases of the ear, larynx, nose and pharynx), **37**, 101–119.

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**, 1–8.

Neff, D. L. (**1995**). "Signal properties that reduce masking by simultaneous random-frequency maskers," J. Acoust. Soc. Am. **98**, 1909–1920.

Oh, E. L., and Lutfi, R. A. (**2000**). "Effect of masker harmonicity on informational masking," J. Acoust. Soc. Am. **108**, 706–709.
Patel, R., and Schell, K. W. (**2008**). "The influence of linguistic content on the Lombard effect," J. Speech Lang. Hear. Res. **51**, 209–220.
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.
Pickett, J. M. (**1956**). "Effects of vocal force on the intelligibility of speech sounds," J. Acoust. Soc. Am. **28**, 902–905.
Pisoni, D. B., Bernacki, R. H., Nusbaum, H. C., and Yuchtman, M. (**1985**). "Some acoustic-phonetic correlates of speech produced in noise," International Conference on Acoustics Speech and Signal Processing, pp. 1581–1584.
Pittman, A. L., and Wiley, T. L. (**2001**). "Recognition of speech produced in noise," J. Speech Lang. Hear. Res. **44**, 487–496.
Simpson, S. A., and Cooke, M. P. (**2005**). "Consonant identification in *N*-talker babble is a nonmonotonic function of *N*," J. Acoust. Soc. Am. **118**, 2775–2778.
Smiljanic, R., and Bradlow, A. R. (**2005**). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. **118**, 1677–1688.
Stanton, B., Jamieson, L., and Allen, G. (**1988**). "Acoustic-phonetic analysis of loud and Lombard speech in simulated cockpit conditions," International Conference on Acoustics Speech and Signal Processing, pp. 331–334.
Steeneken, H. J. M., and Hansen, J. H. L. (**1999**). "Speech under stress conditions: Overview of the effect on speech production and on system performance," International Conference on Acoustics Speech and Signal Processing, pp. 2079–2082.
Stuart, A., Kalinowski, J., Rastatter, M. P., and Lynch, K. (**2002**). "Effect of delayed auditory feedback on normal speakers at two speech rates," J. Acoust. Soc. Am. **111**, 2237–2241.
Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analysis," J. Acoust. Soc. Am. **84**, 917–928.
Tartter, V. C., Gomes, H., and Litwin, E. (**1993**). "Some acoustic effects of listening to noise on speech production," J. Acoust. Soc. Am. **94**, 2437–2440.
Varadarajan, V. S., and Hansen, J. H. L. (**2006**). "Analysis of Lombard effect under different types and levels of noise with application to in-set speaker ID system," International Conference on Spoken Language Processing, pp. 937–940.
Webster, J. C., and Klumpp, R. G. (**1962**). "Effects of ambient noise and nearby talkers on a face-to-face communication task," J. Acoust. Soc. Am. **34**, 936–941.
Womack, B., and Hansen, J. (**1996**). "Classification of speech under stress using target driven features," Speech Commun., special issue on speech under stress, **20**, 131–150.
Xu, Y., Larson,C. R., Bauer, J. J., and Hain, T. C. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.
Young, S., Kershaw, D., Odell, J., Ollason, D., Valtchev, V., and Woodland, P. (**1999**). *The HTK Book 2.2* (Entropic, Cambridge).

Experimental investigation of an American five-string banjo

Laurie A. Stephey and Thomas R. Moore[a)]
Department of Physics, Rollins College, Winter Park, Florida 32789

(Received 25 June 2008; revised 29 July 2008; accepted 14 August 2008)

An experimental investigation of an American five-string banjo is reported. The acoustic spectra of the strings are analyzed and it is found that in many ways banjo strings behave in a manner similar to those on a piano. The terminating impedance presented to the strings at the bridge is reported and it is found to be generally much lower than the characteristic impedance of the strings. It is also shown that, for a properly tuned banjo, the mode shapes of the membrane are symmetric about the center even though there is a large asymmetric force due to the off-center placement of the bridge. Finally, high-speed electronic speckle-pattern interferometry was used to study the motion of the membranic head after a string is plucked, and it is found that the motion is dominated by a rocking motion of the bridge. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2982371]

PACS number(s): 43.75.Gh, 43.40.At [NHF] Pages: 3276–3283

I. INTRODUCTION

The American five-string banjo is an instrument that consists of five metal strings coupled to a circular membrane through a wooden bridge. Its origins are in Africa, but the modern American banjo is the result of changes made to the instrument after its arrival in the United States over 200 years ago. While there are myriad publications on the motion of plucked strings and struck membranes, to our knowledge, there are a very limited number of reports of investigations into the physics of a coupled string and membrane as is the case with the banjo. Recently, the investigation of a simple model of the banjo was reported by Dickey;[1] this was followed by a report of some experimental observations by Rae and Rossing.[2] The recent publication of at least two conference abstracts also indicates a growing interest in the subject.[3,4]

When one attempts to investigate the physics of the banjo, there are two difficulties that are immediately encountered: the difficulty in modeling the instrument due to the complexity of the interactions of the various parts and the lack of experimental data with which to compare any theoretical results. Here we attempt to remedy the latter of these problems in the hope that it will spur efforts to address the former.

In what follows, we present the results of experiments designed to lend some insights into the physics of the banjo. We present measurements of the driving point impedance at the bridge, analysis of the time-averaged and time-resolved acoustic spectra, images of the normal modes of the membrane, and results of studies of the motion of the bridge and membrane after a string has been plucked. These measurements form a foundation for future theoretical investigations that will hopefully result in a complete description of the dynamics of this complex musical instrument.

[a)]Electronic mail: tmoore@rollins.edu

II. EXPERIMENTS

A photograph of the banjo studied in this investigation can be found in Fig. 1. The banjo head is comprised of a membrane stretched over a wooden ring (commonly called a rim) and held in place by a circular metal flange. Connected to the bottom left edge of the rim is a metal handrest and fitted into the rim on the underside of the membrane is a metal tone ring. Some banjos also have a removable wooden back. The rim is attached to a wooden neck and at the top of the neck are tuning pegs to adjust the tension of the strings; the fifth string, however, is shorter than the other four and is adjusted with a tuning peg located at a position near the middle of the neck. Each string passes over a small piece of wood or bone on the neck called a nut, which can be treated as the terminating point of the string. At the base of the head, the strings are attached to the rim by a piece of metal referred to as a tailpiece. A small wooden bridge is placed on the membrane in a position such that a string terminated at the 12th fret has a fundamental frequency one octave higher than that of the same string when terminated at the nut. The strings are stretched over the bridge, which places a downward force on it and ensures continuous contact with the membrane.

All of the experiments reported here were performed on a Gold Tone MC-150R banjo manufactured circa 2005. It has a frosted mylar membrane of thickness 0.21 ± 0.01 mm and density of approximately 1200 kg/m^3. It also has a standard three-footed bridge, shown in Fig. 2, with a length of 80.4 ± 0.1 mm and mass 2.5 ± 0.1 g. It has five metal strings and an arch-top tone ring. The neck is 501 ± 1 mm long, and between 31.9 ± 0.1 mm and 49.6 ± 0.1 mm wide. The radius of the head is 135 ± 1 mm and the rim is 52.3 ± 0.1 mm tall and 20.0 ± 0.1 mm thick. There are 24 screws around the rim that were used to tighten the flange and keep the membrane under tension.

The large number of screws securing the flange around the membrane gives the player the ability to adjust the tension of the head in several places. Most players agree that a symmetric head tension is desirable, and uniform tension can be achieved using several methods, including the use of a

 © 2008 Acoustical Society of America

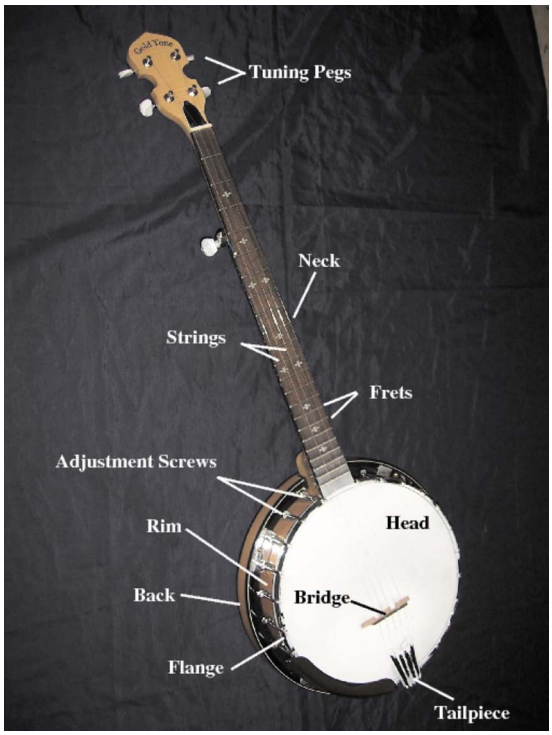


FIG. 1. (Color online) A photograph of the American five-string banjo used in this investigation.

torque wrench or a drum-head tension gauge.[5] Many players, however, tune the head by tapping various points on the membrane and listening to the pitch. They then tighten or loosen the screws to obtain the same pitch at all points on the head. All three methods result in a banjo with approximately symmetric head tension.

In all of the experiments reported here, the wooden back and the metal handrest were removed from the instrument and, unless otherwise noted, the strings were damped with strips of cloth. The head was tuned to approximately 275 Hz (approximately a D4-flat) and a drum head tuner was used to ensure that the membrane remained at uniform tension throughout the experiments. The strings were tuned in the standard G-tuning manner described in Table I.

When the banjo is played in concert, it is often played with the wooden back attached (sometimes called the "resonator"), and there is little doubt that the addition of the back causes a coupling of the air modes and the modes of the membrane. This coupling can shift the resonance frequencies of the membrane, and this effect has been addressed in detail for the case of the tympani;[6] one can reasonably assume that

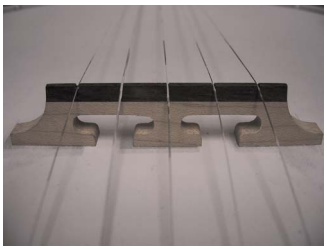


FIG. 2. (Color online) A photograph of the three-footed bridge used in this investigation.

TABLE I. The standard G-tuning of a five string banjo.

String	Note	Fundamental frequency (±0.1 Hz)
1	D	293.7
2	B	246.9
3	G	196.0
4	D	146.8
5	G	392.0

the effect of the cavity will be similar for the banjo. However, while the removal of the back will change the resonance frequencies of the membrane, we have found that the presence of the resonator does not change the deflection shapes of the modes and the configuration investigated in this work allows the experimentalist to address a common situation that can be reproduced with almost any banjo. Additionally, the removal of the back allowed a precise measurement of the driving point impedance as described below.

A. Driving point impedance at the bridge

One of the most important physical parameters for any stringed musical instrument is the impedance presented to the string. As is the case with most stringed instruments, the banjo string is terminated at one end by the nut, which has significantly greater impedance than the characteristic impedance of the string and in most cases can be considered infinite. The end of the string opposite the nut can usually be considered to terminate at the wooden bridge, although it extends approximately 5 cm beyond the bridge to the tailpiece. When a string is plucked, the energy in the string is transferred to the resonating membrane through the bridge. This transfer of energy is governed by the impedance presented to the string at the bridge.

The impedance at the bridge is both a function of the shape and material of the bridge and the characteristics of the membranic head to which it is attached. It is dependent on the driving frequency, and to understand why the banjo produces the sound it does it is critical to know how the impedance presented to the strings varies as a function of frequency. To determine this, we have measured the driving point impedance at the bridge of a fully assembled banjo.

The driving point impedance is the mechanical impedance at the point of excitation and is defined as the driving force divided by the velocity at the point of contact. To measure this quantity, it is essential to obtain simultaneous measurements of both the driving force on the bridge and the velocity of the bridge without significantly altering the instrument. The experimental arrangement used for these measurements is depicted in Fig. 3.

Screws that traditionally are intended to attach the banjo to the wooden back were instead used to attach the banjo to several aluminum posts that were in turn securely mounted on an optical table. The mounting arrangement was such that the neck of the banjo was perpendicular to the table, leaving the area behind the banjo free of any obstruction. The bridge was driven in a sinusoidal manner using an electromagnetic shaker. A load cell designed to measure forces less than 0.01 N was placed between the electromagnetic driver and

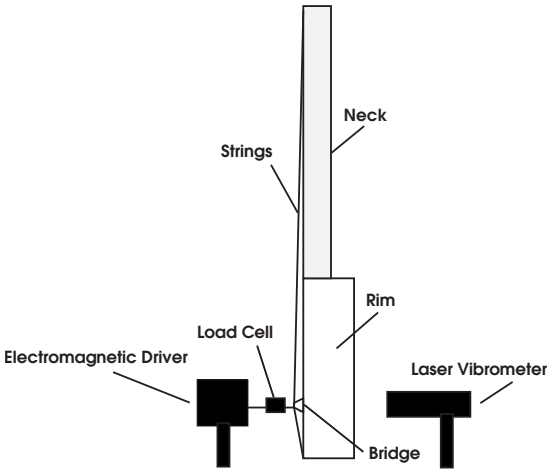


FIG. 3. The experimental arrangement used to measure the impedance presented to the strings at the bridge.

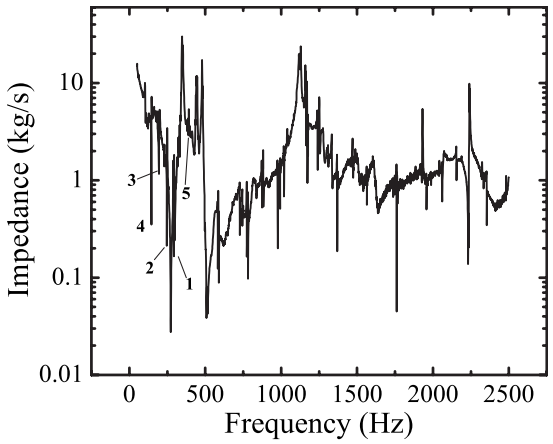


FIG. 4. A plot of the driving point impedance presented to the strings at the bridge vs driving frequency. The numbers indicate the minima at the fundamental frequency of each of the five strings.

the bridge to measure the force exerted on the bridge. The velocity of the resulting motion was measured using a laser vibrometer, which was placed behind the banjo with the beam focused at the point on the membrane directly under the bridge. Measuring the velocity in this manner, rather than the more common method of attaching an accelerometer near the driving point, avoided the possibility of introducing uncertainties into the measurement caused by adding the mass of the accelerometer to the mass of the bridge. Measurements of velocity and force were simultaneously recorded by a computer in real time.

Measurements of impedance were made from 50 to 2500 Hz with 1 Hz resolution. By increasing the driving amplitude of the electromagnetic shaker, three to six measurements of increasing force and velocity were made at each frequency. Because force is linearly related to the velocity through the impedance, the impedance at each frequency was calculated by performing a linear regression on the several measurements of force versus velocity. The uncertainty in the measurements was found by calculating the uncertainty in the slope of this regression and was typically less than 1 kg/s. A plot of the driving point impedance versus frequency is shown in Fig. 4. The data shown in Fig. 4 were taken with the strings left free to vibrate. The effects due to the strings is small but noticeable and the local minima at the fundamental frequencies of the strings are noted in the figure.

B. Acoustic power spectra

The impedance mismatch between the string and the bridge determines how efficiently energy is transferred from the vibrating string to the vibrating membrane. If the kinetic energy was only transferred from the string to the head, it would be a simple matter to model the interaction and determine the dynamics of the string motion. However, because energy can be transferred both to and from the string, accurately accounting for this energy transfer will be one of the more difficult aspects of modeling the banjo. Arguably, the most important result of the two-way energy flow is the length of time that the string continues to vibrate after being plucked. Accurately modeling this aspect of the banjo sound will be imperative for both understanding and electronically reproducing the banjo sound, so it is useful to study the manner in which the sound of a plucked string decays.

We have measured the acoustic power in the first several harmonics of each string while damping the remaining four strings. During these experiments the banjo was mounted in the same way as detailed above. A microphone was placed approximately 40 cm from the front of the banjo head and each string was plucked with a piece of synthetic hair, which was pulled until the hair broke. Each string was plucked in the same manner and the same place, approximately 7 cm above the bridge.

A logical decomposition of the string motion is to assume a linear superposition of motions that are parallel to and perpendicular to the plane of the membrane. The acoustic signal that results from plucking the string in each of these two directions is different, and it is worthwhile to study the sound produced by the banjo in each case. Although the sound produced by plucking each string in each of the two directions is unique to the string, there are a few common characteristics that appear to be important. We only present spectra from the first two strings here because they clearly demonstrate all of these phenomena.

Figure 5 contains the time-averaged power spectra of the first two strings that result from plucking each string vertically and horizontally. Also included are the graphs of the time evolution of the first three harmonics in each case. The decay of many of the resulting harmonics is well described by an exponential function and can be modeled in the usual form $P=P_0e^{-\alpha t}$, where P_0 is the initial power, α is the the decay constant, and t is the time. However, there are also instances of cyclic behavior as well as cases where there are two distinct decay constants for the same string. We will address these issues more fully in Sec. III below.

C. Mode shapes of the membrane

Much can be learned by observing the shapes and frequencies of the normal modes of vibration of a membrane. In the case of the banjo head, the shape of the modes of vibra-

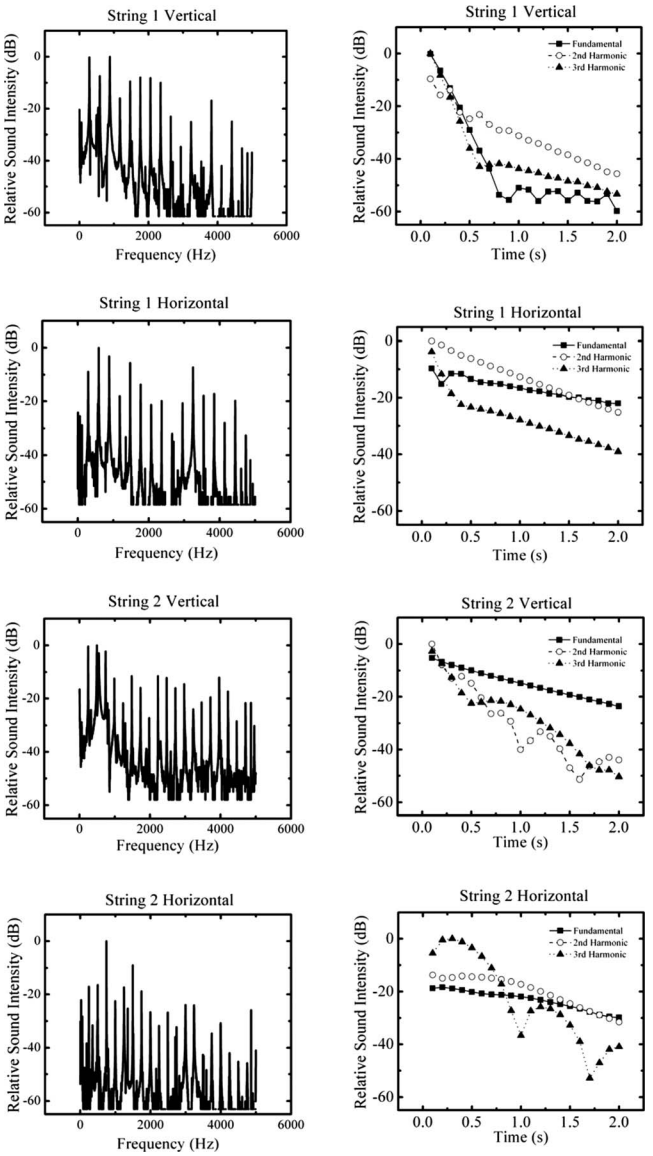


FIG. 5. The time-averaged power spectra resulting from vertically and horizontally plucking each of the first two strings and the plots of the decay of the first three harmonics for each string as a function of time.

tion can lend some insights into the propensity for acoustic short circuiting of the sound, as well as the impact of changing the placement of the bridge on the head. Furthermore, knowledge of the shapes and frequencies of the normal modes can help explain some of the structure observed in the plot of the driving point impedance at the bridge shown in Fig. 4.

The mode shapes of the banjo head were determined using time-averaged electronic speckle-pattern interferometry. The design of this particular interferometer is described elsewhere.[7] The motion of the head was driven acoustically by a speaker connected to a function generator, ensuring that an antinode was not forced at any particular position as can be the case when the vibrations are driven by making contact with the head. The function generator produced a sinusoidal wave with a precision of 0.1 Hz and the speaker was placed approximately 1 m from the front of the banjo. Measurements were made with the strings both damped and undamped; however, it was found that the resonances of the strings did not noticeably affect the resonances of the head. Interferograms that indicate the shapes of nine of the normal modes are shown in Fig. 6 and the frequencies at which these normal modes occur have been cataloged in Table II. In general, the width of the resonances were less than 10 Hz; however, the (0,0) mode could be easily excited over a range of almost 100 Hz.

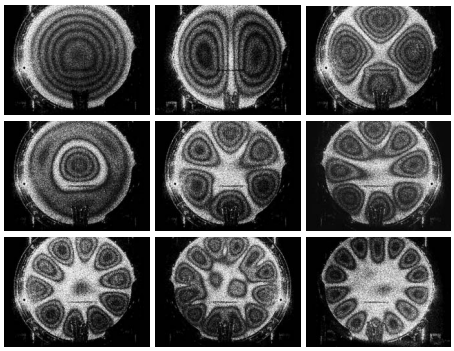


FIG. 6. Time-averaged electronic speckle-pattern interferograms of the membrane motion at nine of the resonant frequencies. Note that the final three interferograms depict deflection shapes with strong components of a normal mode of a circular membrane, but anomalous antinodes are also present near the center. The resonance frequencies for each mode are listed in Table II.

It is difficult to ensure that the patterns shown in interferograms such as those in Fig. 6 are indeed the shapes of the normal modes of the membrane. The best that can often be said is that the interferograms represent deflection shapes; however, the mode shapes of a circular membrane are well established in the literature and it is reasonable to assume that the shapes shown in Fig. 6 are indeed those of the normal modes. Additionally, there exists an orthogonal sister mode to two of the modes shown in Fig. 6 due to a slight asymmetry in the head tension.

D. Motion of the bridge and membrane

Having cataloged the most easily excited modes of the head, it is useful to ask if these modes are preferentially excited by a plucked string. Most other membranaphones are struck, and after the transient motion has subsided, modes

TABLE II. Catalog of the nine most easily excited normal modes of the banjo membrane. For modes with a pronounced doublet, the frequency of the nondegenerate mode is also shown. The asterisk indicates that the interferogram demonstrates a strong component of the mode, but the shape was not exactly that expected of a circularly symmetric membrane.

Mode	Frequency 1 ± 10 Hz
(0,0)	277
(1,0)	550, 580
(2,0)	831, 842
(0,1)	978
(3,0)	1084
(4,0)	1320
(5,0)*	1549
(6,0)*	1771
(7,0)*	1991

that do not have nodes at the point of excitation are left to radiate. In contrast to this usual membranic behavior, the banjo has a membrane that is driven continuously at the three points of bridge contact, sometimes over the course of several seconds. It is therefore important that those wishing to model the complete instrument understand how the bridge actually excites the head motion. To investigate this motion, high-speed electronic speckle-pattern interferometry was employed to study the time-resolved motion of the head.

Electronic speckle-pattern interferometry is a common method for studying the time-averaged motion of vibrating objects, and the usefulness of this technique is demonstrated above. However, in situations where the transient response to excitation is important, the time-averaged deflection shapes do not provide enough information. To study the response of the banjo to a plucked string, an electronic speckle-pattern interferometer was outfitted with a high-speed camera. The result was a system that could resolve displacements of less than 150 nm with a time resolution of 100 μs.

Each string was plucked both vertically and horizontally approximately 9 cm above the top of the rim. The speckle patterns resulting from the motion of the head during and after plucking the string were recorded by the high-speed camera, and the resulting images were processed to yield a series of interferograms. These interferograms represent contour maps of the instantaneous displacement of the membrane and show how the deflection shapes of the banjo head change within the first second after the pluck. An example of the time-resolved behavior of the membrane is shown in Fig. 7 and examples of some instantaneous deflection shapes are shown in Figs. 8–10. The interferograms look the same whether the displacement of the membrane is into or out of the plane; therefore, only one-half of a cycle of the fundamental frequency is shown in Fig. 7.

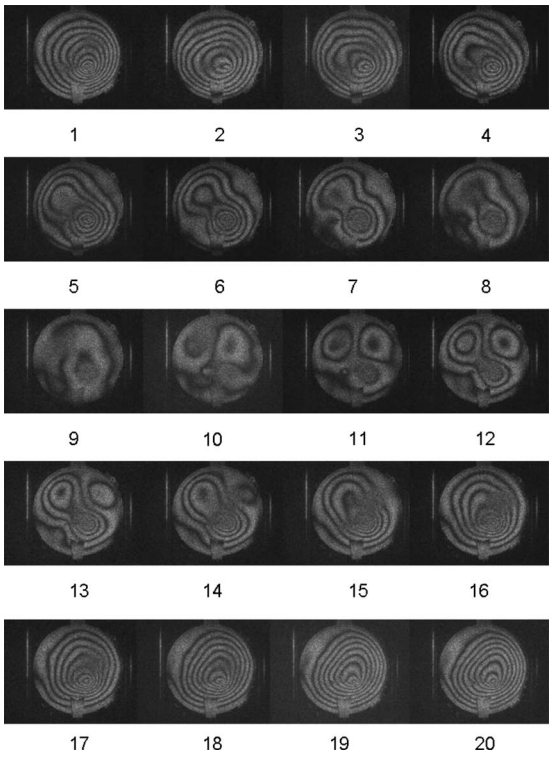


FIG. 7. An example of a series of interferograms of the time evolution of the vibrating head. This motion is the result of the second string plucked vertically and depicts a complete cycle of the second harmonic. Each interferogram depicts the membrane 100 μs after the previous frame. The initial frame occurs 0.4152 s after the pluck. Note that the bridge pivots about its left foot.

III. DISCUSSION OF RESULTS

When taken together, the data presented above provide a basic understanding of the response of the banjo to transient and steady-state excitation. In general, the measurements conform to what one may expect of a string-membrane interface, but there are many interesting effects that deserve to be noted.

It is common to estimate the impedance of the string by assuming it is close to the characteristic impedance of an infinite string, which is given by the square root of the product of the tension and linear density. For a banjo string, this value is on the order of several hundred kg/s; thus the impedance presented to the string at the bridge generally appears to be significantly lower than the impedance of the string itself. Because the characteristic impedance of the string is generally larger than the measured impedance at the bridge, this impedance mismatch should lend itself to efficient energy transfer at most frequencies, but it should also lead to the presence of strong odd harmonics. Since we do not see this evident in the spectrum, it probably indicates that the characteristic impedance of the string is not a good estimate of the actual impedance in this case.

The measurements of the impedance presented to the string at the bridge, which are discussed in Sec. II A and shown in Fig. 4, indicate that there are very few frequencies at which the impedance is above 25 kg/s. However, there are two frequency ranges in which the impedance is relatively large: between 300 and 500 Hz and again between 1000 and 1200 Hz. It is interesting to note that the first peak in impedance roughly corresponds to the range in which most of the fundamental resonances of the strings lie.

It is often more useful to study the admittance, the inverse of impedance, rather than the impedance itself. A graph of the mechanical admittance at the bridge is shown in Fig. 11. As one might expect, the frequencies of several of the

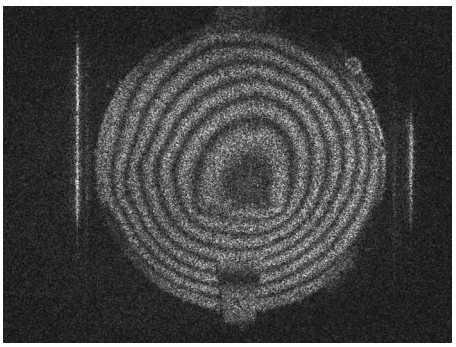


FIG. 8. A single image extracted from a series of time-resolved interferograms. This image shows that vertically plucking the first string results in a strong component of the (0,0) mode. Note that, in this case, the bridge is not the location of an antinode.

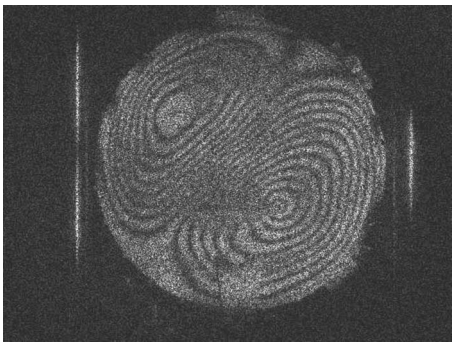


FIG. 9. A single image extracted from a series of time-resolved interferograms. This image shows that vertically plucking the fourth string results in a strong component of the (1,0) mode. Note that the right foot of the bridge is on an antinode but the left foot is not.

admittance maxima are close to those of normal modes of the membrane, and these coincidences have been noted in Fig. 11. We note that these measurements differ considerably from those presented in Ref. 2, and while we have no explanation for this difference, it is probably not due to the structure of the bridge itself since the work reported in Ref. 2 indicates that very little of the structure in the admittance curve can be attributable to the bridge.

The measurements of the acoustic power after a string is plucked are also noteworthy. The power spectra that result from plucking a string horizontally and vertically show that the power generally decreases with increasing harmonic number. This decrease in the power appears not to be a simple function of frequency and is unique to each string. Of particular interest when studying the acoustic power spectra is the placement of the strongest harmonic. The balance of power within the harmonics is a function of the details of the instrument as well as the position on the string at which the pluck occurs, so one must be careful when extending the results to a general conclusion about the sound of any particular banjo; however, the location of the pluck in these experiments was approximately 7 cm above the bridge and was deliberately chosen so as not to fall on a node of any of the first five harmonics of the string. It is also approximately the position that a player normally plucks the string. As one would expect, the direction of the pluck (i.e., horizontally or

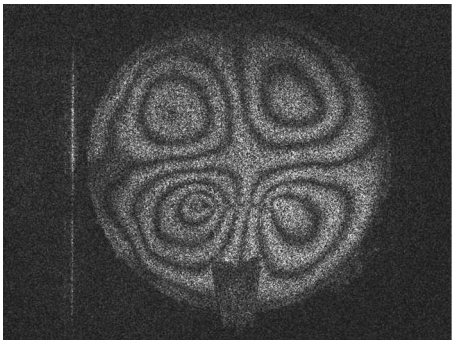


FIG. 10. A single image extracted from a series of time-resolved interferograms. This image shows that vertically plucking the fifth string excited a strong component of the (2,0) mode. Note that each end of the bridge rests on an antinode of the mode shape.

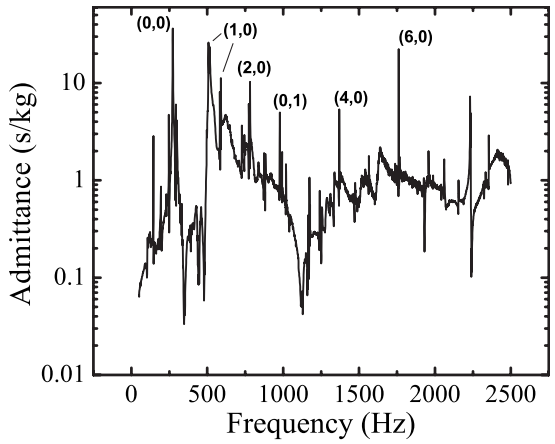


FIG. 11. A plot of the admittance presented to the strings at the banjo bridge vs driving frequency. Admittance maxima which coincide with a normal mode of the membrane are labeled with the mode designation.

vertically) alters the ratio of the power in each harmonic since the motion of the bridge is largely determined by the direction of string motion.

The results shown in Fig. 5 also indicate that the direction of the pluck strongly affects the time it takes for each harmonic to decay. As a rule, horizontally plucked strings decay more slowly than vertically plucked strings. It is also clear that the harmonics decay in one of three ways: exponentially, exponentially with two distinct rates of decay, or with a series of distinct periodic maxima and minima. The direction of the pluck appears to have little impact on which of the functional forms describes the decay of the sound.

The nonexponential decay with evident periodicity is most likely indicative of a slight difference in phase between the sound generated by the vertical and horizontal motions, of string vibration. Since this phase difference varies depending on distance from the instrument, the appearance of the nonexponential decay can change depending on the location of microphone. This also indicates that the horizontal and vertical motions are not necessarily coupled, although evidence for some coupling is discussed below. This lack of strong coupling has also been observed in the high-speed analysis of the deflection shapes of the membrane, where it was determined that plucking the string vertically excited significantly different deflection shapes than plucking the string horizontally.

The harmonics that demonstrate a decay at two distinct rates, such as those evident in the analysis of the first string, most likely correspond to frequencies at which there is a significant difference between the decay rate for the two distinct polarizations of string motion. Thus the change in the slope of the decay occurs as a result of the superpositions of the two motions, each with a significantly different decay rate. Although the plucking mechanism ensured that the initial motion of the string was primarily in the horizontal or vertical direction, high-speed photography indicated that there was usually a transition to elliptical motion, demonstrating that the two directions of motion are not completely decoupled. Indeed, we found it impossible to ensure that the string motion occurs only in one direction using the method of plucking described above.

These observations indicate that the decay characteristics of banjo strings are remarkably similar to those of piano strings. In a study of the decay of piano strings, Weinreich observed both nonexponential periodic decay and twofold exponential decay.[8] The fact that both instruments are comprised of vibrating strings with two independent degrees of motion and coupled oscillators (the head and strings in the banjo and multiple strings in the piano) probably accounts for their similar behavior.

The symmetric shapes of the normal modes of the membrane have been noted previously by Rae and Rossing and the mode shapes shown in Fig. 6 are similar to those reported in Ref. 2. The significant difference between the interferograms in Ref. 2 and those presented here lies primarily in the fact that in producing the interferograms shown in Fig. 6 there was no discrete driving point. That is, in the work reported here, the positions of the antinodes were entirely determined by the physics of the instrument because the vibrations of the membrane were driven acoustically rather than by direct contact with the membrane.

The method used to drive the vibrations is important because there can be significant differences in the orientation of the mode shapes, or even the presence of certain modes, depending on where the membrane is excited. Thus, when the interferograms in Fig. 6 are compared with those in Figs. 8–10, which were produced when the motion is driven by a plucked string, the differences are meaningful. Driving the motion of the head acoustically rather than mechanically also ensures that harmonic distortion of the driving signal is minimized and there is no concern that the excitation is occurring on the nodal line of a mode.

The symmetric deflection shapes evident in Fig. 6 are indicative of the fact that when tuning the banjo head the artist attempts to account for the asymmetric force on the head created by the bridge. Therefore, while the presence of symmetric mode shapes may appear surprising, it is the expected result of a well tuned banjo head. This is fortunate for those attempting to model the complex interaction between the string and membrane; the presence of symmetric mode shapes of the head that are consistent with those expected of a circular membrane indicates that the normal modes of a circular membrane are a logical basis set to use, should one wish to model the head motion using a modal decomposition.

Since the motion of the membrane plays an integral part in producing the final sound, the time-resolved interferograms are some of the most interesting results reported here. Analysis of the interferograms described in Sec. II D shows that the two outer feet of the bridge are extremely important in driving the membrane, whereas the middle foot appears to play a minimal role. This may indicate that accurately modeling the complex shape of the bridge is not necessary to produce a reasonable model of the instrument. Indeed, it is possible that a solid rectangular bridge may be adequate as a first-order approximation for the purpose of modeling the banjo.

The time-resolved interferometry also revealed that it may not be difficult to obtain an intuitive understanding of the bridge motion after a string has been plucked. When a string that terminates on the left side of the bridge was plucked, the rocking motion of the bridge appeared to pivot about the right foot. Conversely, when a string that terminates on the right side of the bridge was plucked, the opposite was true. This pivoting motion can be observed in Fig. 7, where the bridge pivots about the left foot over one complete cycle of the second harmonic. A clearly defined rocking motion of the bridge was present in all but one case: when the third string, located at the center of the bridge, was plucked vertically the bridge did not rock significantly.

As one might also expect, plucking any string excited the membrane into a series of complex deflection shapes, and in most cases the bridge was the location of largest displacement. Most often, the two outer feet of the bridge were antinodes, although in some cases additional antinodes of comparable or greater amplitude were excited elsewhere in the membrane. Generally, the deflection shapes were a complex superposition of the normal modes of a circular membrane; however, in some interesting cases, there was a strong component of a single mode present. Analysis of the head motion as a result of horizontal and vertical plucking, in conjunction with the acoustic data, indicates that when there were strong components of a normal mode of the membrane they were excited by the close proximity of the string harmonics and head resonances, as one might expect.

For example, when the first string was plucked vertically it excited a strong component of the lowest resonance of the head, seen in Fig. 8, due to the close proximity of the fundamental frequency of the string and the resonance frequency of the head. The rapid decay of the fundamental due to this coupling is evident in Fig. 5. Plucking the fourth string vertically resulted in a strong (1,0) component in the motion of the head, as seen in Fig. 9, this is probably due to the close proximity of the resonance of that mode to the third harmonic of the fourth string. Likewise, plucking the fifth string vertically resulted in a strong (2,0) component in the motion of the membrane, seen in Fig. 10, due to the proximity of the frequency of the (2,0) mode to the second harmonic of the fifth string. The interferograms did not indicate any strong isolated components of modes above (2,0).

It is important to note that the interferograms shown in Figs. 8–10 are individual time-resolved frames, not time-averaged interferograms such as those shown in Fig. 6. A superposition of other mode shapes was clearly visible in each time-resolved sequence, but the modes shown in these figures represent a strong component that was visible in isolation at several times during the oscillation. These mode shapes are similar to those shown in Fig. 6, but note the different orientation that results from the fact that the motion was driven by the string rather than an acoustic signal. The orientation of the antinodes in Fig. 9 is particularly interesting because only one of the antinodes corresponds to the position of the a foot of the bridge, but when the membrane is driven freely the two exterior feet of the bridge both fall on an antinode of the motion. Thus, the rocking motion of the bridge pivots about the left foot even though the natural motion of the membrane would have the motion rock symmetrically about the center of the bridge, as is the case shown in Fig. 6. The fact that the bridge motion forces an orientation

of the mode shape that is not normally present when the head is freely vibrating indicates that if the head is not under uniform tension the sound produced by plucking some strings may be altered more significantly than others.

IV. CONCLUSIONS

The banjo is a complex instrument that is just now beginning to be investigated by the community of musical acousticians. Because of the complex interaction between the string, bridge, and membrane, any successful model will require extensive validation. A complete model of this instrument will not only assist in understanding the complex physics responsible for its unique sound, but will also open the door for electronic synthesis of the sound based on first principles. To encourage this process, we have presented the results of several studies of the banjo that lend some insights into the physics and provide a firm basis on which to validate future models. We hope that these results will lead to further theoretical work on this unique instrument.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Grant No. 0551310 and the Rollins College Summer Scholarship Program. The authors wish to thank Gerry Swallowe for the loan of a vibrometer and advice during the course of this research.

[1]J. Dickey, "The structural dynamics of the American five-string banjo," J. Acoust. Soc. Am. **114**, 2598–2966 (2003).

[2]J. Rae and T. Rossing, "The acoustics of the banjo," *Proceedings of ISMA2004*, 2-S1–1, *International Conference on Musical Acoustics* (Acoustical Society of Japan, Nara, Japan, 2004).

[3]J. Dickey, "The banjo in the time-domain," J. Acoust. Soc. Am. **117**, 2539 (2005).

[4]J. Rae and T. Rossing, "Design aspects of the five string banjo," J. Acoust. Soc. Am. **117**, 2590 (2005).

[5]B. Brent, *Earl Scruggs and the 5-String Banjo* (Peer International Corporation, New York, 1968), pp. 17–19.

[6]R. S. Christian, R. E. Davis, and A. Tubis, "Effects of air loading on timpani membrane vibraitons," J. Acoust. Soc. Am. **76**, 1336–1345 (1984).

[7]T. R. Moore and S. A. Zietlow, "Interferometric studies of a piano soundboard," J. Acoust. Soc. Am. **119**, 1783–1793 (2006).

[8]G. Wienreich, "Coupled piano strings," J. Acoust. Soc. Am. **62**, 1474–1484 (1977).

Interaction of reed and acoustic resonator in clarinetlike systems

Fabrice Silva,[a] Jean Kergomard,[b] and Christophe Vergez[c]
Laboratoire de Mécanique et d'Acoustique, UPR CNRS 7051, 13402 Marseille Cedex 20, France

Joël Gilbert[d]
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, 72085 Le Mans Cedex 9, France

(Received 14 January 2008; revised 23 July 2008; accepted 21 August 2008)

Sound emergence in clarinetlike instruments is investigated in terms of instability of the static regime. Various models of reed-bore coupling are considered, from the pioneering work of Wilson and Beavers ["Operating modes of the clarinet," J. Acoust. Soc. Am. **56**, 653–658 (1974)] to more recent modeling including viscothermal bore losses and vena contracta at the reed inlet. The pressure threshold above which these models may oscillate as well as the frequency of oscillation at threshold are calculated. In addition to Wilson and Beavers' previous conclusions concerning the role of the reed damping in the selection of the register the instrument will play on, the influence of the reed motion induced flow is also emphasized, particularly its effect on playing frequencies, contributing to reduce discrepancies between Wilson and Beavers' experimental results and theory, despite discrepancies still remain concerning the pressure threshold. Finally, analytical approximations of the oscillating solution based on Fourier series expansion are obtained in the vicinity of the threshold of oscillation. This allows to emphasize the conditions which determine the nature of the bifurcation (direct or inverse) through which the note may emerge, with therefore important consequences on the musical playing performances.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2988280]

PACS number(s): 43.75.Pq [NHF] Pages: 3284–3295

I. INTRODUCTION

Sound production in reed wind musical instruments is the result of self-sustained reed oscillation. The mechanical oscillator, the reed, acts as a valve which modulates the air flow entering into the instrument by opening and closing a narrow slit defined between the tip of the reed itself and the lay of the mouthpiece. The phenomenon thus belongs to the class of flow-induced vibrations, which has been extensively studied both theoretically and experimentally (see, for example, Ref. 1). Various regimes can occur in such systems: static regime, periodic oscillating regimes, and even complex chaotic behaviors.

A first step to study this kind of oscillator is to analyze the stability of the trivial solution, the equilibrium position of the reed, in order to find a threshold of instability associated with a set of control parameters defining the embouchure (reed, mouthpiece, and player) and the instrument itself. As an output of the threshold analysis, a mouth pressure threshold can be found. It is relevant for the musical playing performance: An estimation of the threshold value is a first evaluation of the ease of playing. In this paper the threshold of instability of the reed of a clarinetlike system is firstly investigated theoretically by revisiting the important pioneering work of Wilson and Beavers.[2] Second, small oscillations beyond the instability threshold are considered.

To answer the question whether the reed equilibrium is stable or not, several theoretical methods are available and have been used to study reed wind instruments. Using the feedback loop analogy, it is known as the free oscillation linear stability problem in a closed loop obtained when the nonlinear component of the loop is linearized around the trivial solution. Then the linear stability of the reed can be studied with respect to each resonance of the input impedance of the resonator in frequency domain (see, for example, Refs. 3–8 for reed instruments, Ref. 9 for lip-reed instruments, and Ref. 10 for a generic type of reeds). Using the dynamical system representation where each resonance of the resonator is described as a simple second-order oscillator in time domain, the linear stability analysis requires to solve eigenvalue problems and to analyze the sign of the eigenvalues real part (see, for example, Ref. 11 for lip-reed instruments and Ref. 12 for vocal folds). The theoretical results can then be compared with experimental ones coming from artificial mouths by playing them as gently as possible (see, for example, Refs. 3, 13, 14, and 2 for reed instruments; Ref. 11 for lip-reed instruments; and Refs. 15, 16, 12, and 17 for vibrating vocal folds). The first attempt to derive theoretically the spectrum of reed instruments beyond the instability threshold is due to Worman.[18] His results were at the origin of several works such as Refs. 19–21.

Despite a rather simple description of physical phenomenon, pioneer theoretical and experimental results concerning mouth pressure and frequency at oscillation threshold

[a]Electronic mail: silva@lma.cnrs-mrs.fr
[b]Electronic mail: kergomard@lma.cnrs-mrs.fr
[c]Electronic mail: vergez@lma.cnrs-mrs.fr
[d]Electronic mail: joel.gilbert@univ-lemans.fr

were obtained by Wilson and Beavers,[2] showing the important role of reed damping determining the clarinetlike or lingual pipe organlike behavior.

The theory of Wilson and Beavers (hereafter denoted WB) is presented and discussed in Sec. II. Then, model improvements from literature are added in the theory in order to try to reduce discrepancies between WB experiments and the theoretical results (Sec. III). Section IV is devoted to a study of small oscillations beyond the instability threshold according to the direct bifurcation behavior of the clarinetlike instruments. Finally perspectives are discussed in Sec. V.

II. WILSON AND BEAVERS THEORY

The model and the principal results of WB are reminded. Particular attention is brought to the numerical estimation of thresholds when the frequency approaches the reed resonance frequency, where some discrepancies with WB results are found. Finally the effect of the reed damping on playability is emphasized.

A. Basic physical model

The physical model used by WB is reviewed with some comments related to more recent literature. It is based on the description of three separate elements: the reed, the bore, and the airflow. The model used here is classical and extremely simplified, but is also proven to be efficient in order to reproduce self-sustained oscillations (for sound synthesis examples, see Refs. 22 and 23). Additional elements will be discussed in the next section.

1. Reed

Based on the fact that reed displacement occurs in the vertical direction mainly without torsion, WB, among many authors, assumed a single degree of freedom motion. Reed-lip-mouthpiece system is thus modeled as a lumped second-order mechanical oscillator with stiffness per unit area K, damping parameter q_r, and natural angular frequency ω_r, driven by the pressure drop $P_m - p(t)$ across the reed, with an inward striking behavior:

$$\frac{d^2y}{dt^2} + q_r\omega_r\frac{dy}{dt} + \omega_r^2(y(t) - y_0) = \frac{\omega_r^2}{K}(p(t) - P_m), \tag{1}$$

$p(t)$, P_m, $y(t)$, and y_0, being the mouthpiece pressure, the blowing pressure, the tip opening [denoted $a(t)$ in WB's paper[2]], and the tip opening without any pressure difference, respectively. P_m is assumed to be constant.

Avanzini *et al.*[24] numerically showed that this lumped model is valid for a small vibration theory where only the interaction between bore resonances and the first flexural mode of the reed is investigated. Measured transfer functions of a reed mounted on a mouthpiece also show a two degree of freedom response.[25]

2. Bore

The behavior of the acoustical resonator is determined by an input impedance relationship between acoustic quantities in the mouthpiece [acoustic pressure $p(t)$ and volume flow $u(t)$, or $P(\omega)$ and $U(\omega)$, respectively, in the frequency domain]. WB assumed for a cylindrical bore representing a simplified clarinet body, an expression given by Backus:[3]

$$\mathcal{Z}_e(\omega) = \frac{P(\omega)}{U(\omega)} = jZ_c\frac{1}{1 - \frac{j}{2Q}}\tan\left(\frac{\omega L}{c}\left(1 - \frac{j}{2Q}\right)\right), \tag{2}$$

where $j^2 = -1$. c, ρ, L, S, and $Z_c = \rho c/S$ are the wave speed in free space, density of air, bore length, bore cross section, and characteristic impedance, respectively. The quality factor Q is assumed by WB to be frequency independent, implying a damping proportional to the frequency.

As it will be seen later, this assumption can be discussed, and improved models will be used because pressure thresholds are strongly influenced by bore losses (which are directly linked to the value of parameter Q).

3. Airflow

As noted by Hirschberg,[26] in the case of clarinetlike instruments, the control of the volume flow by the reed position is due to the existence of a turbulent jet. Indeed, a jet is supposed to form in the mouthpiece (pressure p_{jet}) after flow separation from the walls at the end of the (very short) reed channel. Neglecting the velocity of air flow in the mouth compared to the jet velocity v_{jet} and assuming a downward air flow ($v_{\text{jet}} > 0$), the Bernoulli theorem applied between the mouth and the reed channel[26] leads to

$$P_m = p_{\text{jet}} + \tfrac{1}{2}\rho v_{\text{jet}}^2. \tag{3}$$

Assuming a rectangular aperture of width W and height $y(t)$, the volume flow u across the reed channel can be expressed as follows:

$$u(t) = Wy(t)\sqrt{\frac{2}{\rho}}\sqrt{P_m - p_{\text{jet}}(t)}. \tag{4}$$

Since the cross section of the mouthpiece is large compared to the cross section of the reed channel, it can be assumed that all the kinetic energy of the jet is dissipated through turbulence with no pressure recovery (like in the case of a free jet). Therefore, pressure in the jet is (assuming pressure continuity) the acoustic pressure $p(t)$ imposed by the resonator response to the incoming volume flow u. This model is corroborated by experiments.[27] Similar descriptions are used for double-reed instruments[28] and buzzing lips[29].

B. Characteristic equation and instability threshold

The conditions for which self-sustained oscillations become possible are sought, that is, the minimum value of blowing pressure required for the static regime to be unstable is investigated for a given configuration of the experiment (ω_r, q_r, and L being fixed). Common methods of linear stability analysis[30] are used in this study, and solutions having time dependence $\exp(j\omega t)$ are sought. Cancellation of the imaginary part of ω corresponds to an oscillation that is neither damped nor amplified: it characterizes the instability threshold of static regime. Attention is drawn to the fact that

this quantity may differ from the oscillation threshold depending on the nature of the bifurcation, studied in Sec. IV B. As a language abuse *oscillation threshold* is often used instead of *instability threshold*.

Assuming small vibrations around equilibrium state (mean values of y and p are $y_0 - P_m/K$ and 0, respectively), the volume flow relationship (4) is linearized. Dimensionless quantities are introduced here: θ, $\mathcal{Y}_e$, and D are the dimensionless frequency, input admittance, and the reed transfer function, respectively.

$$\theta = \frac{\omega}{\omega_r}, \quad \mathcal{Y}_e(\theta) = \frac{Z_c}{\mathcal{Z}_e(\theta)} \text{ and } D(\theta) = \frac{1}{1 + jq_r\theta - \theta^2}. \tag{5}$$

There are two dimensionless control parameters:

$$\gamma = \frac{P_m}{Ky_0} \text{ and } \zeta = Z_c W \sqrt{\frac{2y_0}{K\rho}}. \tag{6}$$

γ is the ratio between mouth pressure and the pressure required to completely close the reed channel in static regime, while ζ mainly depends on mouthpiece construction and lip stress on the reed and is linked to the maximum flow through the reed channel (ζ equals quantity 2β in Ref. 2). Linearization of Eq. (4) leads to the so-called characteristic equation:

$$\mathcal{Y}_e(\theta) = \zeta\sqrt{\gamma}\left\{D(\theta) - \frac{1-\gamma}{2\gamma}\right\}, \tag{7}$$

which can be split into real and imaginary parts:

$$\mathrm{Im}(\mathcal{Y}_e(\theta)) = \zeta\sqrt{\gamma}\,\mathrm{Im}(D(\theta)), \tag{8}$$

$$\mathrm{Re}(\mathcal{Y}_e(\theta)) = \zeta\sqrt{\gamma}\left(\mathrm{Re}(D(\theta)) - \frac{1-\gamma}{2\gamma}\right). \tag{9}$$

At last, a dimensionless length $k_rL = \omega_r L/c$ is introduced.

C. Numerical techniques

The unknowns θ, $\gamma \in \mathbb{R}^+$ satisfying Eq. (7) are numerically determined for a range of bore lengths, parameters (q_r, ζ, ω_r) being set. They correspond to frequency and mouth pressure at instability threshold of the static regime. When various solutions exist for a given configuration due to the interaction of the reed resonance with the several bore resonances, the threshold observed experimentally by increasing the blowing pressure is the one having the minimum value of γ.

The characteristic equation is transcendental and may have an infinite number of solutions. Zero finding is done using the Powell hybrid method,[31] which combines the advantages of both Newton method and scaled gradient one. A continuation technique is adopted to provide an initial value to the algorithm: the first resolution is done for very high values of L ($k_rL \simeq 30$), i.e., for a nearly nonresonant reed (i.e., with neither mass nor damping, hereafter denoted the massless reed case), where $f \simeq (2n-1)c/4L$ and $\gamma \simeq 1/3$ (with $n \in \mathbb{N}^*$). Bore length is then progressively decreased and zero finding for a given value of L is initialized with the pair (θ, γ) solution of the previous solving (bore slightly longer). Depending on the initialization of the first resolution ($k_rL = 30$), it is possible to explore the branches associated with the successive resonances of the bore. When reed resonance and bore antiresonance get closer to each other ($k_rL \to n\pi (n \in \mathbb{N})$), fast variation of the pressure threshold requires adjustment of the step size.

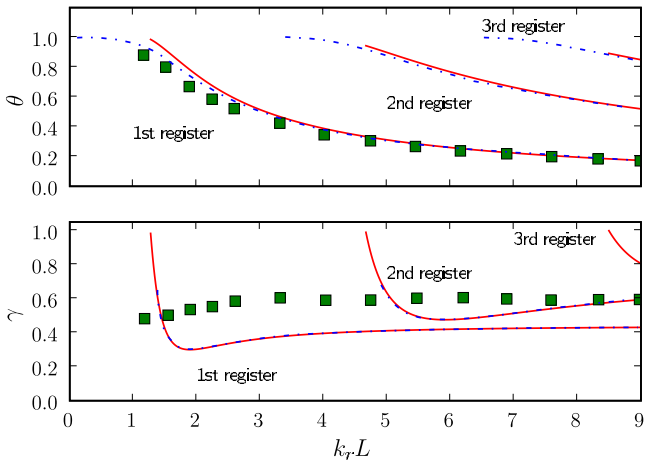


FIG. 1. (Color online) Dimensionless threshold frequencies (top) and pressure (bottom) for a strongly damped reed: $q_r = 0.4$, $f_r = 750$ Hz, and $\beta = 0.065$. Results (dashed lines) and measurements (squares) from Ref. 2; our numerical results (solid lines).

D. Results

Two kinds of behavior can be distinguished. For strongly damped reeds (Fig. 1), the threshold frequencies (dashed lines, result from Ref. 2) always lie either near the reed resonance ($\theta = 1$) or near the first impedance peak frequency of the pipe [hyperbola $\theta = \pi/(2k_rL)$, not represented in the figure for readability], corresponding to the first register of the instrument. When the length L decreases, pressure threshold gradually reduces from values assumed for the massless reed model to a minimum point for $k_rL \sim \pi/2$, and then strongly increases as the pipe becomes shorter. When increasing γ from 0, the loss of stability of the static regime may give rise to an oscillating solution which always corresponds to the first register since the instability thresholds of the higher registers occur for higher values of mouth pressure. On the contrary, considering now slightly damped reeds (Fig. 2), emerging oscillations can occur near higher pipe resonances. Indeed, for certain ranges of L, the pressure threshold associated with one particular higher-order register is lower than the pressure required to drive the air column in the first register. This lowest pressure threshold is associated with the acoustic mode, the natural frequency of which being the closest to the reed resonance.

These results show the influence of reed damping on selecting

- a clarinetlike behavior with heavily damped cane reed: preference is given to the *chalumeau* register, i.e., the lowest register, or
- a lingual organ pipe behavior with very lightly damped metallic reed: tuning is performed by adjusting reed vibrating length by means of a wire adjusted on reed, see Ref. 32 for further details.

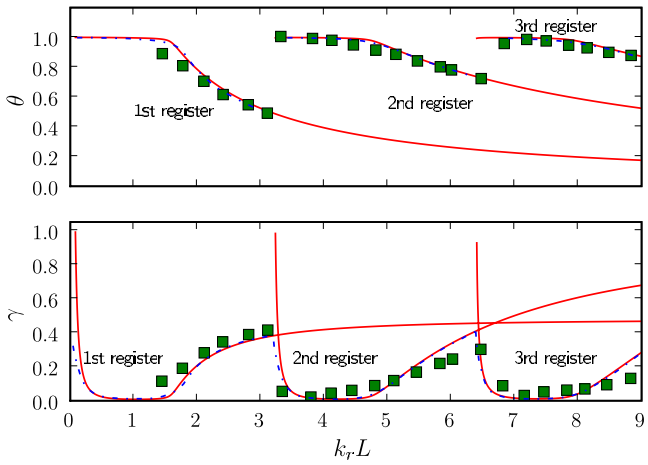


FIG. 2. (Color online) Dimensionless threshold frequencies and pressure for a lightly damped reed: q_r=0.008, f_r=700 Hz, and β=0.05. Results (dashed lines) and measurements (squares) from Ref. 2; our numerical results (solid lines).

This noticeable conclusion of Wilson and Beavers' paper gainsays von Helmholtz,[33] who stated that the differences of behavior are linked to the mass of the reed.

E. Discussion

Despite the important results obtained by WB, discrepancies concerning numerical results and model limits can be pointed out. In a validation phase of our numerical algorithms, our results (see Figs. 1 and 2) were compared to their ones (the data being extracted from their article), solving exactly the same equation. Differences appear between numerical results concerning threshold frequencies, approaching reed resonance when the length decreases. Investigations pointed out model limits the authors do not take into account: linearization of flow relationship is valid only while reed channel is not closed at rest, i.e., when $P_m < Ky_0$. Using the linear form for higher values of P_m would be meaningless, even for free reed aerophones: the opening function (linked to the reed displacement) taking part in flow calculation can never be negative. For instruments for which reed beats against the mouthpiece, reed channel is completely closed and then sustained oscillations cannot occur for bore lengths where WB theory predicts pressure threshold above static beating reed pressure by extending linearization beyond model limits.

The fact that the dimensionless mouth pressure γ remains lower than 1 implies the existence of a maximum threshold frequency for the oscillation. We assume that the maximum value $\theta_{\max}$ is associated with the maximum value of γ (i.e., 1). No theoretical proof is given here, but this assumption is observed in all numerical results obtained. Then the solution $\theta_{\max}$ is given by

$$\mathcal{Y}_e(\theta_{\max}) = \zeta D(\theta_{\max}), \tag{10}$$

which leads to relation

$$1 - \theta_{\max}^2 = \zeta \,\mathrm{Re}(\mathcal{Z}_e(\theta_{\max})) > 0 \tag{11}$$

as the bore is a passive system. As a conclusion, it is not possible to play sharper than a frequency slightly flatter than the reed-lip-mouthpiece system resonance frequency, the deviation to this latter being linked to the losses in the bore.

F. Minimum pressure threshold: Improved playability for interacting resonances

For each value of q_r, there exists one or more ranges of bore lengths where playability is greatly improved. Indeed, pressure threshold curves show a minimum for a certain value of k_rL, denoting an increased easiness to produce the note corresponding to this length. Associating a clarinet mouthpiece with a trombone slide, informal experiments confirm that it is easier to produce some notes than other ones. Analytical approximated expressions for the minimum have been investigated. Under the assumption that this minimal value is obtained for an emerging frequency located close to the reed resonance, and therefore is mainly determined by reed damping, bore losses can be ignored $\mathrm{Re}(\mathcal{Y}_e(\omega))=0$, Eq. (9), leading thus to

$$\gamma = \frac{1}{1 + 2\,\mathrm{Re}(D(\theta))}. \tag{12}$$

In the massless reed model ($D(\theta)=1$), threshold pressure is equal to 1/3 and frequencies correspond to frequencies for which the imaginary part of the bore input impedance vanishes, which is consistent with results already published.[34] The minimum pressure threshold occurs at a maximum of $\mathrm{Re}(D(\theta))$:

$$\mathrm{Re}(D(\theta)) = \frac{1 - \theta^2}{(1 - \theta^2)^2 + (q_r\theta)^2}, \tag{13}$$

obtained for $\theta = \sqrt{1 - q_r}$ (which is consistent with the approximation $\theta \simeq 1$), thus,

$$\gamma_0 = \frac{q_r(2 - q_r)}{2 + q_r(2 - q_r)}, \tag{14}$$

nearly proportional to q_r for lightly damped reeds.

In order to evaluate the effect of bore losses on this minimum pressure threshold, a one-mode resonator with losses is now considered:

$$\mathcal{Y}_e(\omega) = Y_n\left(1 + jQ_n\left(\frac{\omega}{\omega_n} - \frac{\omega_n}{\omega}\right)\right), \tag{15}$$

where Y_n is the minimum amplitude of the admittance and Q_n is the quality factor; Eqs. (8) and (9) become

$$Y_n + \zeta\frac{1 - \gamma}{2\sqrt{\gamma}} = \zeta\sqrt{\gamma}\frac{1 - \theta^2}{(1 - \theta^2)^2 + (q_r\theta)^2}, \tag{16}$$

$$Y_nQ_n\left(\frac{\theta}{\theta_n} - \frac{\theta_n}{\theta}\right) = -\zeta\sqrt{\gamma}\frac{q_r\theta}{(1 - \theta^2)^2 + (q_r\theta)^2}. \tag{17}$$

Ignoring the variation of Y_n with the length of the bore, the derivation of Eq. (16) with respect to k_rL leads to a minimum value of the function $\gamma = f(k_rL)$ for $\theta_{\min}^2 = 1 - q_r$ and $\gamma_{\min}$ solution of equation

$$\frac{Y_n}{\zeta\sqrt{\gamma_{\min}}} + \frac{1-\gamma_{\min}}{2\gamma_{\min}} = \frac{1}{q_r(2-q_r)}, \tag{18}$$

which the first-order solution is given by

$$\gamma_{\min} \simeq \gamma_0\left(1 + 2\frac{Y_n}{\zeta}\sqrt{\gamma_0}\right) \tag{19}$$

obtained for

$$\omega_n = \omega_r\left(1 - \frac{q_r}{2} + \frac{1}{2Q_n} + \frac{\zeta}{2Y_nQ_n\sqrt{q_r}}\right). \tag{20}$$

For an open/closed cylinder $\omega_n=(2n-1)\pi c/2L$, the result is

$$(k_rL)_{\min} \simeq (2n-1)\frac{\pi}{2}\left(1 + \frac{q_r}{2} - \frac{1}{2Q_n} - \frac{\zeta}{2Y_nQ_n\sqrt{q_r}}\right). \tag{21}$$

Typical values $Y_n=1/25$, $\zeta=0.4$, and $q_r=0.4$ lead to an increase in $\gamma_{\min}$ relative to γ_0 of about 8%, confirming the preponderant effect of reed damping on the minimum pressure threshold.

In order to understand how coupling acoustical and mechanical resonances could reduce pressure threshold over a wider bore length range, the neighborhood of the previously mentioned minimum has been studied. Using again a single acoustical mode for the calculation, derivation of a parabolic approximation was possible. Writing $\gamma=\gamma_{\min}(1+\varepsilon^2)$, $\theta^2=\theta_{\min}^2+\delta$, and $\omega_n=(\omega_n)_{\min}(1+\upsilon)$, with ε, δ, and υ small quantities, the Taylor expansion of Eqs. (16) and (17) with respect to these values leads to the next relationships:

$$\varepsilon^2 \sim \delta^2/(2q_r^2), \tag{22}$$

$$2q_r^2Y_nQ_n(\delta-2\upsilon) = -\delta\zeta\sqrt{q_r}. \tag{23}$$

Finally, near the minimum pressure threshold, dependence on bore length is given by

$$\gamma = \gamma_{\min}\left(1 + \frac{2q_r}{\left(q_r^{3/2} + \frac{\zeta}{2Y_nQ_n}\right)^2}\left(\frac{k_rL - k_rL_{\min}}{k_rL_{\min}}\right)^2\right). \tag{24}$$

In a first approximation, the aperture of the approximated parabola, thus the width of the range for which oscillation threshold is lowered, is mainly controlled by the musician embouchure, i.e., by reed damping and lip stress on the reed. This means that, thanks to its embouchure, the player can expect an easier production of tones for certain notes.

For a lossy cylindrical open/closed bore, modal expansion of input impedance gives $Y_nQ_n=\omega_nL/2c=(2n-1)\pi/4$ so that bore losses do not seem to have a great influence on playing facility, at least when considering minimal blowing pressure $\gamma_{\min}$ (Q_n does not appear alone in first-order calculation). On the contrary, they are essential for understanding the extinction threshold phenomenon,[35] i.e., when the reed is held motionlessly against the lay.

III. MODEL IMPROVEMENTS

Last four decades have been fruitful in physical modeling of musical instruments, especially for single reed instruments. Pipes have been the focus of a great number of studies since Benade[36] as well as the description of peculiarities of the flow (Backus,[3] Hirschberg,[26] and Dalmont *et al.*[27]). The aim here is to try to reduce discrepancies between WB experiments and theory, based on some of those investigations which look relevant to the study of oscillation threshold. In a real clarinet-player system, the vocal tract may have an effect on the oscillation threshold (see, e.g., Ref. 37). However, the use of porous material in the WB apparatus suggests that acoustic interaction between the chamber and the mouthpiece has been canceled or at least strongly reduced. Therefore we do not discuss the effect of the upstream part.

A. Viscothermal losses model and vena contracta

It should be noticed that there is a gap in pressure threshold values between experiment and theory in WB article. This occurs even for long bores when reed dynamical behavior should not deviate from the ideal spring model (because of an emerging frequency much smaller than ω_r). For that case, Kergomard *et al.*[21] provided an approximated formula taking into account both reed dynamics and bore losses that can be extended to

$$\gamma \simeq \frac{1-\theta_n^2}{3-\theta_n^2} + \frac{2\,\mathrm{Re}(\mathcal{Y}_e(\theta_n))}{3\sqrt{3}\zeta}, \quad \theta_n = (2n-1)\frac{\pi}{2k_rL}. \tag{25}$$

θ_n corresponding to the nth resonance frequency of the bore. In comparison with ideal model (lossless bore and massless reed, i.e., $\gamma=1/3$), additional corrective terms are considered in Eq. (25), one lowering pressure threshold due to the collaboration of the resonant reed, the other one requiring higher blowing pressure due to dissipation in the bore. According to this approximated expression, pressure thresholds depend on the mouthpiece parameter and on bore dissipation at playing frequency.

Now focus is done on using realistic values of ζ and $\mathcal{Y}_e(\theta)$, assuming acoustic losses in clarinetlike bore to be due mainly to viscothermal dissipation (Ref. 36). Other kinds of losses such as nonlinearity localized at the open end of a tube are negligible since study is done at oscillation threshold, i.e., for very small amplitude oscillations. Simpler model ("Raman's model") has been recently investigated by Dalmont *et al.*[35] and led to satisfactory oscillation threshold of the fundamental register of the clarinet when operating frequencies are much lower than the reed natural frequency (the reed being considered as an ideal spring). In the present study, the magnitudes of the higher-order impedance peaks need to be correctly estimated so that neither the loss model of Raman nor that of Backus[3] may be realistic enough. Pressure thresholds would be inaccurate when higher-order modes of bore oscillate first.

The standard formula for the input impedance will be hence considered:

$$\mathcal{Z}_e(\omega) = jZ_c\tan(kL) \quad \text{with } jk = \frac{j\omega}{c} + \eta\sqrt{\frac{j\omega}{c}} \tag{26}$$

where η is a coefficient quantifying the viscothermal boundary layers, equals to 0.0421 in mks units for a 7 mm radius cylinder. This model introduces dissipation and dispersion

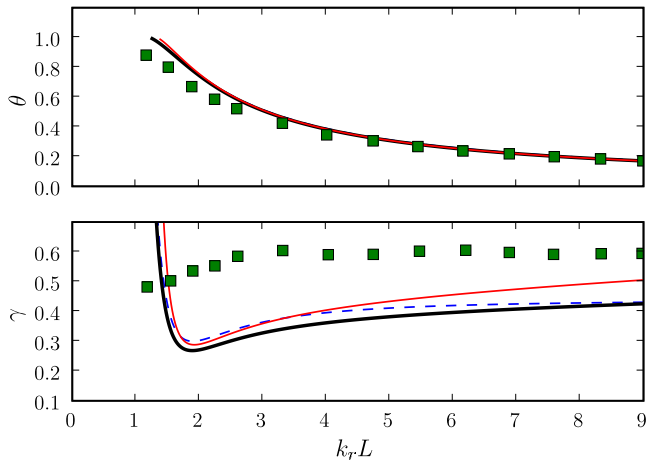


FIG. 3. (Color online) Results with the model considering vena contracta phenomenon (thin line) and viscothermal losses (thick line) (same conditions as in Fig. 1). WB results are also recalled (dashed line).

and leads to a zero valued impedance at zero frequency, which is still consistent with the linearization of flow relationship with a zero mean value of acoustic pressure. In this expression, viscothermal effects are ignored in the characteristic impedance (see Ref. 34). Equation (26) leads to peak magnitudes decreasing with length L, whereas they are not sensitive to L in Backus empirical expression. A direct consequence is that the pressure threshold increases as the resonator lengthens.

Another effect may occur and modify pressure threshold. Hirschberg[26] and Hirschberg *et al.*[38] brought to attention on vena contracta phenomenon: due to sharpness of edges, flow separation may result in the formation of a free jet in reed channel, this contraction effect resulting in a jet cross-sectional area smaller than the reed channel opening. Recent investigations[39] applying the lattice Boltzmann method to the reed channel confirm previous experiments.[40] The assumption of constant vena contracta is valid in some specific cases, for short channel geometry, but there is no proof yet that this effect is significant for a real clarinet mouthpiece. Here small vibrations of the reed near oscillation threshold are considered, which may induce little influence of the flow unsteadiness on the measured pressure threshold. So it is possible[26] to include vena contracta phenomenon by reducing the area of the reed channel $Wy(t)$ by a coefficient approximately 0.6, i.e., by multiplying ζ by this coefficient.

Numerical investigations point out (see Fig. 3) that taking into account realistic losses and vena contracta phenomenon reduces discrepancies in pressure threshold, especially for high values of bore length and for a strongly damped reed, i.e. when reed dynamics has a small influence on oscillation threshold. Nevertheless pressure values experimentally obtained by WB still remain quite higher than the ones corresponding to the modified model, while frequencies are unaltered.

An attempt to explain the discrepancies in threshold measurements will be provided in Sec. IV.

B. Reed motion induced flow

The influence of the reed is not limited to its resonance. Since Nederveen[41] and Thompson,[42] it is proved that the flow entering through the reed opening is divided in one part exciting the resonator and another part induced by the reed motion. In fact, the vibration of the surface of the reed produces an additional oscillating flow. Thus the entering flow U can be written as

$$U(\omega) = \mathcal{Y}_e(\omega)P(\omega) + S_r(j\omega Y(\omega)), \tag{27}$$

where S_r is the effective area of the vibrating reed related to the tip displacement $y(t)$. Alternately, a length Δl can be associated with the fictitious volume where the reed swings. Dalmont *et al.*[14] reported typical values of 10 mm for a clarinet. Nederveen[41] linked Δl to reed strength (or *hardness*): Δl may approximately vary from 6 mm (strong reeds) to 9 mm (softer reeds). These values being small compared to clarinet dimensions, the reed motion induced flow can be considered through a mere length correction in common work, but its influence on the interaction between acoustic resonator and reed is not negligible on the threshold frequency, as it is studied now.

In a first step, this effect is considered separately, all losses being ignored ($q_r=0$ and $\eta=0$). Equation (27) coupled to Eq. (7) leads to the following system:

$$-\mathrm{Im}(\mathcal{Y}_e) = k_r \Delta l \frac{\theta}{1-\theta^2}, \tag{28}$$

$$\frac{1}{1-\theta^2} - \frac{1-\gamma}{2\gamma} = 0 \Leftrightarrow \gamma = \frac{1-\theta^2}{3-\theta^2}. \tag{29}$$

As seen previously, several frequency solutions θ exist for different pressure thresholds γ. Examination of Eq. (29) as a function $\gamma=f(\theta)$ (for $\theta<1$) reveals that the solution having the lowest threshold is the one being the closest to the reed frequency.

Approximations can be derived in some situations. When playing close to a bore resonance frequency $\theta_n \ll 1$, the right-hand term in Eq. (28) is small, so that, with $\mathcal{Y}_e=-j\cot(\theta k_r L)$, the reed motion induced flow acts merely as a length correction:

$$\Delta l \frac{1}{1-\theta_n^2} \quad \text{where } \theta_n = \frac{(2n-1)\pi}{2k_r L}. \tag{30}$$

This approximation is valid when the considered bore frequency θ_n remains smaller than unity. An equivalent approximated length correction can be derived for the effect of reed damping on frequencies from Eq. (8):

$$\Delta l_q \simeq \zeta \frac{q_r}{\sqrt{3}k_r}. \tag{31}$$

A trial and error procedure has been performed to adjust the oscillation frequency. Very few iterations were needed to exhibit a value for the reed motion induced flow ($\Delta l \simeq 12$ mm) higher than for the reed damping ($\Delta l_q \simeq 2$ mm) in the conditions of Figs. 1 and 4. When acoustical and mechanical resonances are very close ($\theta=1-\varepsilon$ and $\theta_n=1-\varepsilon_n$), a second-order expression can be deduced:

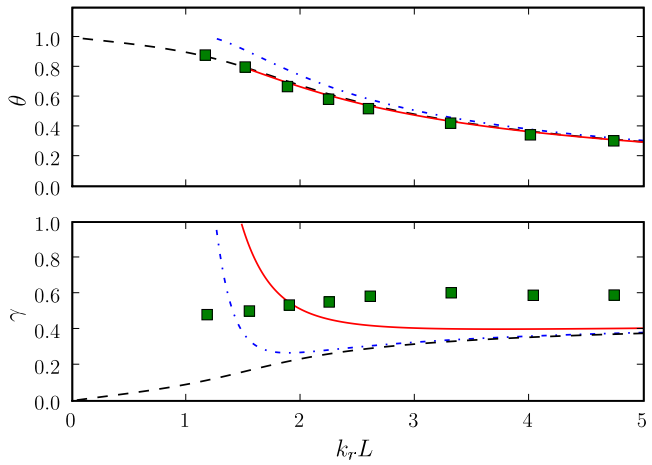


FIG. 4. (Color online) Comparing reed motion induced flow effect and reed damping effect (Δl=12 mm and other conditions as in Fig. 1): reed motion induced flow only (dashed lines), reed damping only (dash-dotted lines), and both effects (plain lines). WB experimental results are recalled (squares).

$$\varepsilon = \frac{\varepsilon_n}{2}\left(1 + \sqrt{1 + 2\frac{\Delta l}{L\varepsilon_n^2}}\right), \tag{32}$$

the apparition of a square root being typical of mode coupling, making the achievement of analytical expressions difficult. Then, when the bore length decreases enough so that one of its resonances increases above the reed one ($\theta_n > 1$), the oscillation frequency approaches the reed one until the intersection point disappears for $k_rL = n\pi$ (see Fig. 5). Near the reed resonance, first-order approximations can be derived:

$$\theta \simeq 1 - \tfrac{1}{2}k_r\Delta l \tan(k_rL), \tag{33}$$

$$\gamma \simeq \tfrac{1}{2}k_r\Delta l \tan(k_rL). \tag{34}$$

These expressions are valid if $\tan(k_rL) \gtrsim 0$, i.e., $k_rL \gtrsim n\pi$. According to Eq. (29), when oscillation frequency approaches f_r, the pressure threshold decreases to zero contrary to what happens when considering the reed damping effect.

Figures 4 and 6 show a numerical comparison of the respective effects of reed motion induced flow and reed damping. The reed motion induced flow with Δl=12 mm in Fig. 4 and 5 mm in Fig. 6 adjusts the frequency deviation for

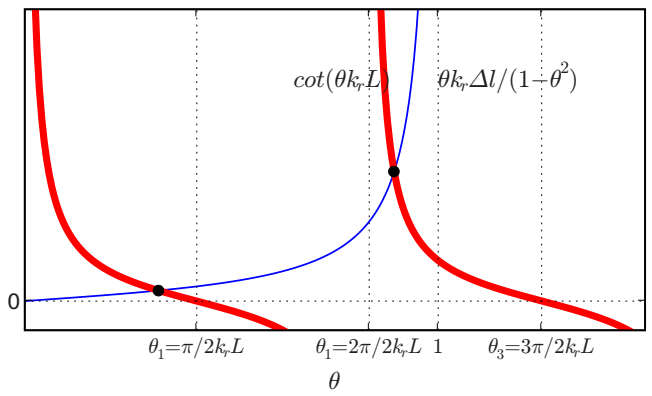


FIG. 5. (Color online) Graphical representation of Eq. (28) giving oscillation frequency at threshold: left-hand term- Im($\mathcal{Y}$) (thick lines), right-hand term $k_r\Delta l\theta/(1-\theta^2)$ (thin line), and solutions (markers).

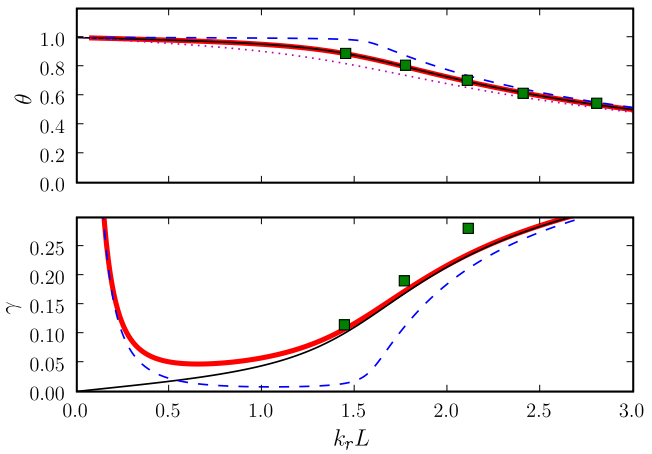


FIG. 6. (Color online) Comparing reed motion induced flow effect and reed damping effect (condition as in Fig. 2): reed motion induced flow only (Δl=5 mm: plain thin lines, Δl=12 mm: dotted line in upper graph only), reed damping only (dashed thin lines), and both effects (plain thick lines). WB experimental results are recalled (squares).

both heavily and slightly damped reeds, even when approaching reed resonance, and is preponderant compared to the damping effect. An example of a mistuned length correction parameter (Δl=12 mm) is exhibited in Fig. 6, revealing a mismatch of oscillation frequency. This latter seems to be mainly controlled by the reed motion induced flow effect when approaching the reed resonance frequency. Classical approaches (see, e.g., Ref. 41) for the calculation of playing frequencies, ignoring the flow due to pressure drop, and searching for eigenfrequencies of the passive system including the bore and the reed only are thus justified. It can be noticed that Eq. (28) was already given by Weber[43] in the early 19th century (see page 216), assuming a reed area equal to the cylindrical tube section. This theory, used by several authors (see, e.g., Ref. 32), was discussed by Von Helmholtz[33] and Bouasse,[44] especially concerning the lack of explanation on the production of self-sustained oscillations. On the contrary, threshold pressure curves exhibit that both damping and additional flow have influence on the pressure required for the reed to oscillate. So a combination of the phenomena has to be taken into account, none of them being negligible in the considered domain.

IV. GOING BEYOND INSTABILITY THRESHOLD

A. Linear stability analysis with modal decomposition

Analysis of the instability threshold can be performed using complex frequency formalism. For a given configuration of the whole system *bore-reed-musician* (L, S, ω_r, q_r, γ, and ζ being set), its complex eigenfrequencies $s_n = j\omega_n - \alpha_n$ can be determined. The imaginary part of s_n corresponds to the frequency, the real part α_n being the damping of this mode, for the coupled linearized system close to the static

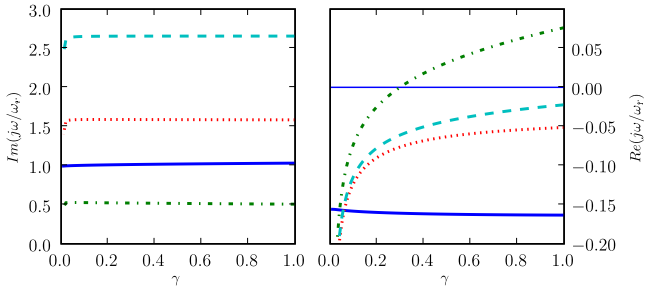


FIG. 7. (Color online) Evolution of the complex eigenfrequencies as a function of mouth pressure γ. $r=7$ mm, $\omega_r=2\pi\times1000$ rad/s, $q_r=0.3$, $L=16$ cm ($\theta_1=0.53$ and $k_rL=2.96$), and $\zeta=0.2$.

equilibrium state. Classically, when the mouth pressure is below the oscillation threshold, all eigendampings α_n are positive, the static regime being stable. An oscillation may appear when at least one mode of the whole system becomes unstable, i.e., when at least one of the α_n becomes negative. Looking for instability threshold can be done by varying a bifurcation parameter (either L, γ, or ζ) and examining the real part of computed eigenfrequencies. Two examples are shown in Figs. 7 and 8. It is noticeable that the frequencies evolve only slightly with the bifurcation parameter γ and are close to eigenfrequencies of either the bore or the reed ($\mathrm{Im}(j\omega/\omega_r)\simeq1$). For the first example, the static regime becomes unstable for $\gamma\simeq0.28$ and a frequency near the first resonance of the bore (dot-dashed curves). Other acoustic resonances (their frequencies being odd multiples of the first one) and reed resonance have higher oscillation threshold and remain damped for this configuration. For a longer tube (Fig. 8), instability appears for $\gamma\simeq0.3$ at a frequency located near the third resonance of the bore [dotted line at $\mathrm{Im}(j\omega/\omega_r)\simeq0.8\sim3\times0.27$], the first resonance becoming unstable for a larger mouth pressure. The frequency close to the reed resonance (solid curves) still remains damped.

Calculations may be simplified and accelerated by using a modal decomposition of the bore impedance $\mathcal{Z}_e(\omega)$: this allows for the characteristic equation to be written as a polynomial expression of $j\omega$, and optimized algorithms for polynomial root finding can be used. Modal expansion considers the N first acoustic resonances of the bore:

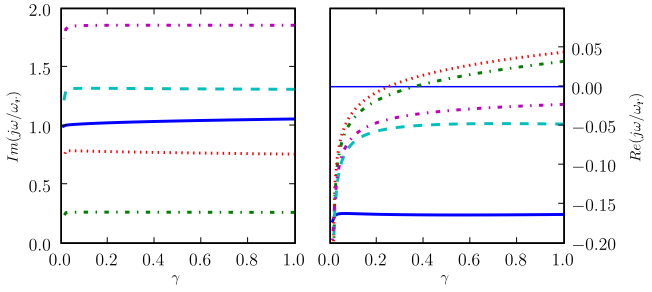


FIG. 8. (Color online) Evolution of the complex eigenfrequencies as a function of γ. $r=7$ mm, $\omega_r=2\pi\times1000$ rad/s, $q_r=0.3$, $L=32$ cm ($\theta_1=0.26$ and $k_rL=5.91$), and $\zeta=0.2$.

TABLE I. Comparison of pressure threshold and oscillation frequency calculated using complex frequency formalism (index by CF) and Wilson and Beavers' method (WB) for $r=7$ mm, $\omega_r=2\pi\times750$ rad/s, $q_r=0.4$, and $\zeta=0.13$.

k_rL	θ_{CF}	θ_{WB}	$\Delta\theta$	γ_{CF}	γ_{WB}	$\Delta\gamma$
8.5	0.185	0.186	0.5%	0.43	0.43	0.1%
2	0.747	0.747	0.1%	0.30	0.30	0.6%
1	1.022	1.024	0.2%	3.86	3.82	1.0%
0.81	1.033	1.034	0.1%	8.77	8.71	0.7%

$$\frac{\mathcal{Z}_e(\omega)}{\mathcal{Z}_c}=j\tan\left(\frac{\omega L}{c}-j\alpha(\omega)L\right) \tag{35}$$

$$\frac{\mathcal{Z}_e(\omega)}{\mathcal{Z}_c}\simeq\frac{2c}{L}\sum_{n=1}^{N}\frac{j\omega}{\omega_n^2+jq_n\omega\omega_n+(j\omega)^2}, \tag{36}$$

where modal coefficients ω_n and q_n can be deduced either from measured input impedance or from analytical expression [Eq. (35)], assuming $\alpha(\omega)$ to be a slowly varying function of the frequency.

Comparison between direct calculation of oscillation threshold γ_{th} using WB method and estimation using modal decomposition and complex eigenfrequencies computing has been done. Whereas the difficulties for the first method arise due to the transcendental characteristic equation, the second one requires calculation of eigenvalues for various values of the mouth pressure γ, using an iterative search of the instability threshold. The number of modes taken into account has been chosen such that resolution of γ_{th} is less than 0.01, which is also the tolerance used for the iterative search. An example is given in Table I for a heavily damped and strong reed. It shows very good agreement between the two methods for both γ and θ. This validation allows the use of the complex frequency approach, which results in an efficient algorithm that can be easily applied to more complex resonators whenever a modal description is available.

The writing of the characteristic equation for a single acoustic mode exhibits the behavior of the coupled oscillators:

$$\left[\omega_n^2+j\omega\left(q_n\omega_n+\frac{c}{L}\zeta\frac{1-\gamma}{\sqrt{\gamma}}\right)-\omega^2\right]\times\left[\omega_r^2+jq_r\omega\omega_r-\omega^2\right]$$
$$=j\omega\frac{2c}{L}\omega_r^2\left(\zeta\sqrt{\gamma}+j\omega\frac{\Delta l}{c}\right). \tag{37}$$

Coupling realized by the flow in the reed channel modifies the damping of the acoustic mode: in addition to the usual term (corresponding to viscothermal losses and eventually radiation), damping is increased by a quantity related to mouth pressure and stress on the reed. This may be regarded to as a resistive acoustic behavior at the bore entrance.

Assuming that a linearized model is still relevant during the growth of oscillations (before the amplitude saturation mechanism appears), this approach can be extended to investigate the transient response of the coupled system. Characterizing the degree of instability of the system by $\sigma=\min_n\alpha_n$, the slope of the curve $\sigma=f(\gamma)$ gives an informa-

tion about the instability degree of the system when the mouth pressure is slightly higher than the oscillation threshold. A great slope would correspond to a very unstable configuration and a quick growth of oscillation, whereas nearly constant curve would lead to a small amplification coefficient and slowly rising vibrations and then to longer transient attack before stabilization of the magnitude of oscillations. Links between the computed eigenfrequencies of the coupled system presented here and the transient behavior have still to be investigated.

B. Mouth pressure required to obtain a given (small) amplitude

The previous sections of the paper deal with the stability of the static regime, looking for the condition to make a bifurcation possible. Some developments concerning the existence of oscillating regime above the threshold are derived now. Neither stability of oscillation nor tone deviation issue will be discussed here.

Grand *et al.*[20] suggested the introduction of the limited Fourier series of pressure in the massless reed case. The technique is the harmonic balance applied to oscillations of small amplitudes. Calculations are done hereafter by taking into account the reed dynamics in the volume flow relationship, which does not appear in the mentioned paper. Fourier series of the volume flow depends on Fourier components of signal $P_m - p(t)$ and $y(t)$. Assuming steady state oscillations with angular frequency ω, the dimensionless signals are written as

$$p(t) = \sum_{n\neq 0} p_n e^{nj\omega t}, \quad u(t) = u_0 + \sum_{n\neq 0} \mathcal{Y}_n p_n e^{nj\omega t}, \tag{38}$$

$$\frac{y(t)}{y_0} = (1-\gamma) + \sum_{n\neq 0} D_n p_n e^{nj\omega t}, \tag{39}$$

where $\mathcal{Y}_n = \mathcal{Y}_e(n\omega)$ and $D_n = D(n\omega)$ are the values of dimensionless bore admittance and reed transfer function for angular frequency $n\omega$. The volume flow relationship is rewritten as

$$u^2(t) = \zeta^2 (y(t)/y_0)^2 (\gamma - p(t)). \tag{40}$$

Sustained oscillations of very small amplitude are studied, assuming that p_1 is a nonvanishing coefficient considered as a first-order quantity. Notations $\mathbf{CE}_n$ and $\mathbf{F}\binom{n}{m}$ are introduced:

$$\mathbf{CE}_n = \mathcal{Y}_n/(\zeta\sqrt{\gamma}) + \frac{1-\gamma}{2\gamma} - D_n, \tag{41}$$

$$\mathbf{F}\binom{m}{n} = \mathbf{F}\binom{n}{m} = D_n D_m - \frac{1-\gamma}{\gamma}(D_n + D_m) - \frac{\mathcal{Y}_m \mathcal{Y}_m}{\zeta^2 \gamma}. \tag{42}$$

Cancellation of $\mathbf{CE}_n$ for given ω and γ means that the characteristic equation (7) is solved for $n\omega$ and γ. Expanding Eq. (40) leads to

$$0 = \left[\frac{u_0^2}{\zeta^2\gamma} - (1-\gamma)^2\right] + 2(1-\gamma)\sum_{n\neq 0}\left[\frac{u_0 \mathcal{Y}_n}{\zeta^2\gamma(1-\gamma)} - D_n + \frac{1-\gamma}{2\gamma}\right] p_n e^{nj\omega t} - \sum_{n,m\neq 0} \mathbf{F}\binom{n}{m} p_n p_m e^{(n+m)j\omega t} + \frac{1}{\gamma}\sum_{n,m,q\neq 0} D_n D_m p_n p_m p_q e^{(n+m+q)j\omega t}. \tag{43}$$

It is here assumed that p_n is of order $|n|$ (with $p_{-n} = p_n^*$) (see, e.g., Ref. 20). The continuous component of the volume flow is calculated up to order 2:

$$\frac{u_0^2}{\zeta^2\gamma} = (1-\gamma)^2 + \sum_{n\neq 0} \mathbf{F}\binom{+n}{-n} |p_n|^2 - \frac{1}{\gamma}\sum_{n,m,n+m\neq 0} D_n D_m p_n p_m p_{n+m}^* \simeq (1-\gamma)^2 + 2\mathbf{F}\binom{+1}{-1}|p_1|^2 + o(p_1^2) \tag{44}$$

so that, considering u_0 to be real,

$$u_0 \simeq \zeta\sqrt{\gamma}(1-\gamma)\left[1 + \frac{|p_1|^2 \mathbf{F}\binom{+1}{-1}}{(1-\gamma)^2}\right] + o(p_1^2). \tag{45}$$

From Eq. (43), the frequency $(N\omega)$ is extracted for $N \geq 1$:

$$0 = 2(1-\gamma)\left[D_N - \frac{1-\gamma}{2\gamma} - \frac{u_0 \mathcal{Y}_n}{\zeta^2\gamma(1-\gamma)}\right] p_N + \sum_{n\neq 0} \mathbf{F}\binom{n}{N-n} p_n p_{N-n} - \frac{1}{\gamma}\sum_{n,m\neq 0} D_n D_m p_n p_m p_{N-n-m}. \tag{46}$$

For $N \geq 2$, Taylor series expansion up to order N is applied: in the first sum, only terms corresponding to $0 \leq n \leq N$ contribute at order N. while, in the second one, the terms to consider are the ones for which $0 < n < N$ and $0 < m < N-n$. The component p_N can be deduced from the sequence $(p_n)_{0<n<N}$:

$$p_N = \frac{1}{2(1-\gamma)\mathbf{CE}_N}\left[\sum_{0<n<N} \mathbf{F}\binom{n}{N-n} p_n p_{N-n} - \frac{1}{\gamma}\sum_{\substack{0<n<N \\ 0<m<N-n}} D_n D_m p_n p_m p_{N-n-m}\right] + o(p_1^N). \tag{47}$$

As an example, for $N=2$, the second sum being empty:

$$p_2 \simeq \frac{\mathbf{F}\binom{1}{1} p_1^2}{2(1-\gamma)\mathbf{CE}_2} + o(p_1^2). \tag{48}$$

As expected and in agreement with the so-called "Worman rule,"[18] higher components appear to be higher-order quantities: order 2 for p_2, 3 for p_3, 4 for p_4, etc.

Focus is now given to fundamental frequency. Calculations are done up to order 3:

$$0 = 2(1-\gamma)\left[D_1 - \frac{1-\gamma}{2\gamma} - \frac{u_0\mathcal{Y}_1}{\zeta^2\gamma(1-\gamma)}\right]p_1 + \sum_{n\neq 0,1}\mathbf{F}\binom{n}{1-n}p_n p_{1-n} - \frac{1}{\gamma}\sum_{1-n-m,n,m\neq 0} D_n D_m p_n p_m p_{1-n-m}$$

$$\simeq 2(1-\gamma)\left[D_1 - \frac{1-\gamma}{2\gamma} - \frac{u_0\mathcal{Y}_1}{\zeta^2\gamma(1-\gamma)}\right]p_1 + 2\mathbf{F}\binom{2}{-1}p_2 p_1^* - \frac{1}{\gamma}D_1 p_1|p_1|^2(D_1 + 2D_1^*). \tag{49}$$

Replacing u_0 and p_2 gives

$$|p_1|^2 \simeq \frac{2(1-\gamma)^2\mathbf{CE}_1}{\dfrac{\mathbf{F}\binom{1}{1}\mathbf{F}\binom{+2}{-1}}{\mathbf{CE}_2} - \dfrac{2\mathcal{Y}_1}{\zeta\sqrt{\gamma}}\mathbf{F}\binom{+1}{-1} - \dfrac{1-\gamma}{\gamma}(D_1^2 + 2|D_1|^2)}. \tag{50}$$

The limitation to the first harmonic method ignoring the influence of $p_{n\geqslant 2}$ on the amplitude $|p_1|$ leads to

$$|p_1|^2 \simeq \frac{2(1-\gamma)^2\mathbf{CE}_1}{-\dfrac{2\mathcal{Y}_1}{\zeta\sqrt{\gamma}}\mathbf{F}\binom{+1}{-1} - \dfrac{1-\gamma}{\gamma}(D_1^2 + 2|D_1|^2)}. \tag{51}$$

In their paper, Grand *et al.*[20] stated that the procedure used in a first simplified case can be applied to a model including reed dynamics provided that $\mathcal{Z}_e(\omega)$ is replaced by $\mathcal{Z}_e(\omega)D(\omega)$. Conclusion is not so straightforward as reed dynamics also interferes with the volume flow relationship contributing to a more complex expression for the first component amplitude. According to their work, D_n terms would only occur with $\mathcal{Y}_n$ in the expression of $|p_1|^2$ which is not the case.

In Fig. 9 are shown the bifurcation diagrams for a heavily damped reed. From the flattest tone (k_rL=9.00) to the sharpest computed (k_rL=1.25), oscillations with very small amplitudes are possible as the diagrams show Hopf bifurcation that is supercritical, i.e., direct, for all the computed cases. On the other hand, for cases where oscillation frequency is closed to the reed one (Fig. 10), a subcritical Hopf bifurcation occurs for a range of bore length, with a very small pressure threshold linked to the interaction of reed resonance with the second bore resonance: the bifurcation is there inverse.

In accordance with Grand *et al.*,[20] it appears that the bifurcation is not always direct. There exist configurations for which computed bifurcation diagram shows a subcritical pitchfork, i.e., small oscillations for values of pressure below static regime instability threshold. The boundary between the two cases is not trivial to explore analytically, requiring the mathematical study of Eq. (50). Alternately numerical exploration of parameter space can lead to some partial observations. One of them is that the bifurcation seems to be direct when the oscillation frequency is close to one of the bore resonances.

The limit of calculations derived here needs to be emphasized. When $|p_1|$ tends to zero, the oscillation frequency is the one for which the characteristic equation is solved ($\mathbf{CE}_1$=0). Higher components (the second one here at the

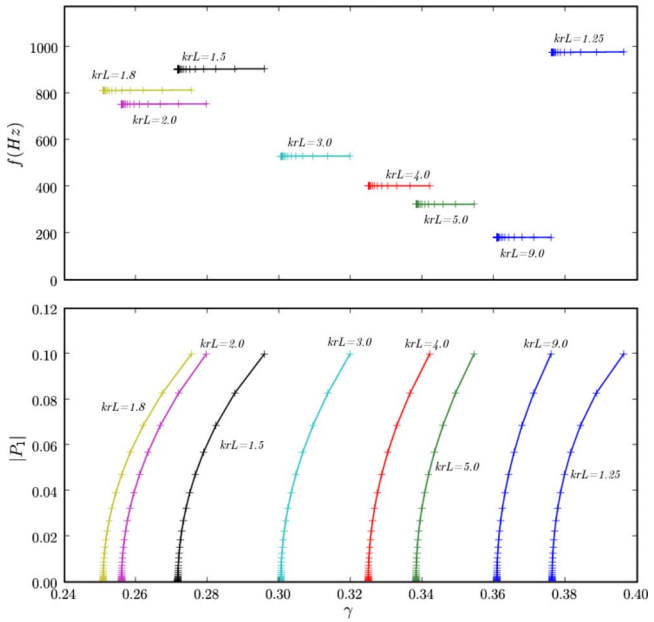


FIG. 9. (Color online) Comparison of $|p_1|$ computed with the harmonic balance for small oscillations for various values of k_rL: from thin to thick lines, k_rL=9,5,4,3,2,1.8,1.5,1.25 (Δl=0, q_r=0.4, and f_r=1050 Hz).

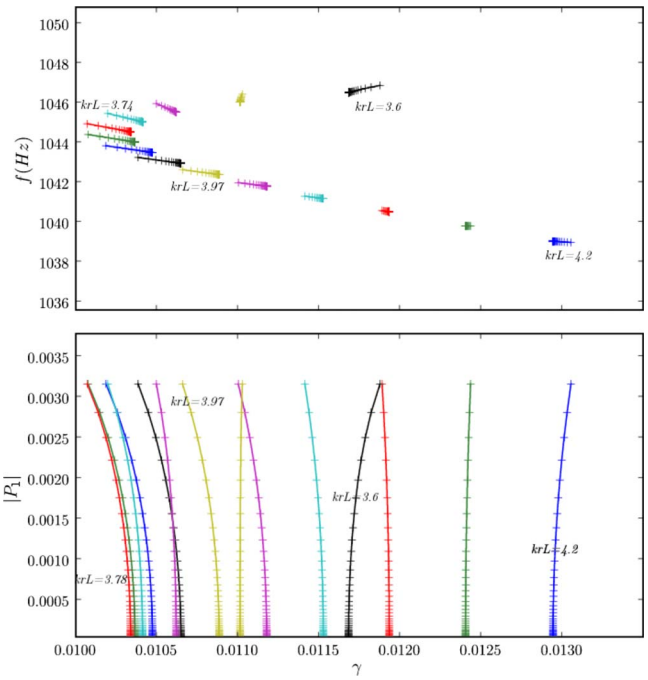


FIG. 10. (Color online) Comparison of $|p_1|$ computed with the harmonic balance for small oscillations for 14 values of k_rL regularly spaced between 3.6 and 4.20 with a lightly damped reed (ΔL=0, q_r=0.01, and f_r =1050 Hz).

first-order approximation) have a non-negligible influence only if $\mathbf{CE}_2$ becomes small too when approaching oscillation threshold, i.e., if the characteristic equation accepts the solutions ω and 2ω for two close values of mouth pressure: $\mathbf{CE}_1$ and $\mathbf{CE}_2$ are small simultaneously. Inverse bifurcation may occur in the degenerate case of simultaneously destabilization of static regime for a frequency and its octave. This consideration can be extended to the more general case of $\mathbf{CE}_1$ and $\mathbf{CE}_N$ $(N \geq 2)$ canceling an equal mouth pressure value.

As a conclusion, the nature of the bifurcation depends on the roots of the characteristic equation (7): this generalizes the results of Grand *et al.*[20]

V. CONCLUSION AND PERSPECTIVES

Two components of the volume velocity at the input of the resonator act on the oscillation threshold values: the first one is due to the pressure drop between mouth and mouthpiece, while the second one is due to the reed movement. Roughly speaking, the first one has a mainly resistive effect, either passive or active, the second one having a mainly reactive effect. This remark can be related to the behavior of the threshold pressures and frequencies. Concerning the frequencies, the second effect is preponderant, at least for the cases studied by Wilson and Beavers, and it can justify the historical method due to Weber[43] regarding the playing frequencies as eigenfrequencies of a passive resonator. Concerning the pressure thresholds, the flow due to the pressure drop is essential, and the model based on Bernoulli equation used by Wilson and Beavers[2] seems to be satisfactory, but taking into account the flow due to reed movement is necessary, and improves the results of the authors, mainly for lightly damped reeds.

Discrepancies between Wilson and Beavers' experimental results and numerical ones still remain. Can this be due to the nature of the bifurcation at threshold? For a clarinetlike functioning, i.e., for strongly damped reeds, numerical calculation of small oscillations proves that it is supercritical, confirming the works ignoring reed dynamics, while it is not sure for lightly damped reeds. Further experimental investigations are planned to conclude on the nature of the bifurcation on real clarinets. Another unknown concerns the acoustic impedance of the air supply device and of the mouth cavity.

Concerning the possible generalization of this work, other kinds of either resonators or reeds should be studied, as, e.g., in Ref. 8. Some authors of the present paper recently obtained simplified theoretical results for an outward striking reed by using a similar method of investigation based on a single degree of freedom (DOF) reed model.[45] Calculations with several DOFs for the reed remain to do, especially for lip-reed instruments (see Ref. 11), even if recent works dedicated to vocal folds have been done (see Ref. 12). Finally, the study of transient sounds can be made easier with a good knowledge of the linearized functioning of reed instruments.

ACKNOWLEDGMENTS

The study presented in this paper was led with the support of the French National Research Agency (ANR) within the CONSONNES project. The authors are grateful to A. Hirschberg, Ph. Guillemain, and B. Ricaud for useful discussion.

[1] R. D. Blevins, *Flow-Induced Vibration*, updated ed. (Krieger, Malabar, FL, 2001).
[2] T. A. Wilson and G. S. Beavers, "Operating modes of the clarinet," J. Acoust. Soc. Am. **56**, 653–658 (1974).
[3] J. Backus, "Small-vibration theory of the clarinet," J. Acoust. Soc. Am. **35**, 305–313 (1963).
[4] A. H. Benade and D. J. Gans, "Sound production in wind instruments," Ann. N.Y. Acad. Sci. **155**, 247–263 (1968).
[5] Y. M. Chang, "Reed stability," J. Fluids Struct. **8**, 771–783 (1994).
[6] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer, Berlin, 1991), pp. 345–494.
[7] J. Saneyoshi, H. Teramura, and S. Yoshikawa, "Feedback oscillations in reed woodwind and brasswind instruments," Acustica **62**, 194–210 (1987).
[8] A. Z. Tarnopolsky, N. H. Fletcher, and J. C. S. Lai, "Oscillating reed valves—An experimental study," J. Acoust. Soc. Am. **108**, 400–406 (2000).
[9] S. J. Elliott and J. M. Bowsher, "Regeneration in brass wind instruments," J. Sound Vib. **83**, 181–217 (1982).
[10] N. H. Fletcher, "Autonomous vibration of simple pressure-controlled valves in gas flows," J. Acoust. Soc. Am. **93**, 2172–2180 (1993).
[11] J. S. Cullen, J. Gilbert, and M. Campbell, "Brass instruments: Linear stability analysis and experiments with an artificial mouth," Acta Acust. **86**, 704–724 (2000).
[12] N. Ruty, "Modèles d'interactions fluide parois dans le conduit vocal. Applications aux voies et aux pathologies (Fluid wall interaction modelling in the vocal tract. Applications to voice and pathologies)," Ph.D. thesis, INPG Grenoble, 2007.
[13] J.-P. Dalmont and C. Frappé, "Oscillation and extinction thresholds of the clarinet: Comparison of analytical results and experiments," J. Acoust. Soc. Am. **122**, 1173–1179 (2007).
[14] J.-P. Dalmont, B. Gazengel, J. Gilbert, and J. Kergomard, "Some aspects of tuning and clean intonation in reed instruments," Appl. Acoust. **46**, 19–60 (1995).
[15] R. W. Chan and I. R. Titze, "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," J. Acoust. Soc. Am. **119**, 2351–2362 (2006).
[16] I. Lopez, A. Hirschberg, A. van Hirtum, N. Ruty, and X. Pelorson, "Physical modeling of buzzing artificial lips: The effect of acoustical feedback," Acta Acust. **92**, 1047–1059 (2006).
[17] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," J. Acoust. Soc. Am. **97**, 1874–1884 (1995).
[18] W. E. Worman, "Self-sustained nonlinear oscillations of medium amplitude in clarinet-like systems," Ph.D. thesis, Case Western Reserve University, (1971).
[19] A. H. Benade and S. N. Kouzoupis, "The clarinet spectrum: Theory and experiment," J. Acoust. Soc. Am. **83**, 292–304 (1988).
[20] N. Grand, J. Gilbert, and F. Laloë, "Oscillation threshold of woodwind instruments," Acta Acust. **83**, 137–151 (1997) URL: http://hal.archives-ouvertes.fr/hal-00001412/fr/ (Last viewed 2007).
[21] J. Kergomard, S. Ollivier, and J. Gilbert, "Calculation of the spectrum of self-sustained oscillators using a variable truncation method: Application to cylindrical reed instruments," Acta Acust. **86**, 685–703 (2000).
[22] P. Guillemain, J. Kergomard, and T. Voinier, "Real-time synthesis of clarinet-like instruments using digital impedance models," J. Acoust. Soc. Am. **118**, 483–494 (2005).
[23] R. T. Schumacher, "Ab initio calculations of the oscillation of a clarinet," Acustica **48**, 71–85 (1981).
[24] F. Avanzini and M. van Walstijn, "Modelling the mechanical response of the reed-mouthpiece-lip system of a clarinet. Part I. A one-dimensional distributed model," Acta Acust. **90**, 537–547 (2004).
[25] B. Gazengel, T. Guimezanes, J.-P. Dalmont, J.-B. Doc, S. Fagart, and Y. Léveillé, "Experimental investigation of the influence of the mechanical characteristics of the lip on the vibrations of the single reed," Proceedings

of International Symposium on Musical Acoustics, Barcelona, Spain, 2007.
[26]A. Hirschberg, *Mechanics of Musical Instruments*, CISM Courses and Lectures No. 355 (Springer, Wien, 1995), Chap. 7, pp. 291–369.
[27]J.-P. Dalmont, J. Gilbert, and S. Ollivier, "Nonlinear characteristics of single-reed instruments: Quasistatic volume flow and reed opening measurements," J. Acoust. Soc. Am. **114**, 2253–2262 (2003).
[28]A. Almeida, C. Vergez, and R. Caussé, "Quasi-static non-linear characteristics of double-reed instruments," J. Acoust. Soc. Am. **121**, 536–546 (2007).
[29]J. Chick, S. Bromage, M. Campbell, S. Stevenson, and J. Gilbert, "Motion of the brass player's lips during extreme loud playing," *Actes du 8ème Congrès Français d'Acoustique (Proceedings of the French Conference in Acoustics)* (SFA, Tours, 2006).
[30]J. K. Hale and H. Kocak, *Dynamics and Bifurcations* (Springer-Verlag, New York, 1991).
[31]M. J. D. Powell, *Numerical Methods for Nonlinear Algebraic Equations* (Gordon and Breach, London, 1970), Chap. VII, pp. 87–144.
[32]A. Miklos, J. Angster, S. Pitsch, and T. D. Rossing, "Interaction of reed and resonator by sound generation in a reed organ pipe," J. Acoust. Soc. Am. **119**, 3121–3129 (2006).
[33]H. von Helmholtz, *On the Sensation of Tone* (Dover, New York, 1954), Appendix VII, pp. 388–394.
[34]J. Kergomard, *Mechanics of Musical Instruments*, CISM Courses and Lectures No. 355 (Springer, Wien, 1995), Chap. 6, pp. 229–290.
[35]J.-P. Dalmont, J. Gilbert, J. Kergomard, and S. Ollivier, "An analytical prediction of the oscillation and extinction thresholds of a clarinet," J. Acoust. Soc. Am. **118**, 3294–3305 (2005).
[36]A. H. Benade, *Fundamentals of Musical Acoustics* (Dover, New York, 1976), Chap. XXI, pp. 430–464.
[37]C. Fritz, "Clarinette et clarinettiste : Influence du conduit vocal sur la production du son (The clarinet and the player: Role of the vocal tract on the sound production)," Ph.D. thesis, Université Paris 6 and University of New South Wales, 2004.
[38]A. Hirschberg, R. W. A. van de Laar, J.-P. Marrou-Maurières, A. P. J. Wijnands, J. H. Dane, S. G. Kruijswijk, and A. J. M. Houtsma, "A quasi-stationary model of air flow in the reed channel of single-reed woodwind instruments," Acustica **70**, 146–154 (1990).
[39]A. R. da Silva, G. P. Scavone, and M. van Walstijn, "Numerical simulations of fluid-structure interactions in single-reed mouthpieces," J. Acoust. Soc. Am. **122**, 1798–1809 (2007).
[40]J. van Zon, A. Hirschberg, J. Gilbert, and A. P. J. Wijnands, "Flow through the reed channel of a single reed music instrument," J. Phys. (Paris) **51**(C2), 821–824 (1990).
[41]C. J. Nederveen, *Acoustical Aspects of Woodwind Instruments* (Northern Illinois, Deklab, IL, 1998).
[42]S. C. Thompson, "The effect of the reed resonance on woodwind tone production," J. Acoust. Soc. Am. **66**, 1299–1307 (1979).
[43]W. Weber, "Theorie der zungenpfeifen (Theory of reed pipes)," Ann. Phys. Chem. **93**, 193–246 (1829).
[44]H. Bouasse, *Instruments à Vent (Wind Instruments)*, first published in 1930 ed. (Librairie Scientifique et Technique Albert Blanchard, Paris, 1986), Vol. **1**.
[45]F. Silva, J. Kergomard, and C. Vergez, "Oscillation thresholds for "striking outwards" reeds coupled to a resonator," Proceedings of the International Symposium on Musical Acoustics, Barcelona, Spain, 2007.

Influence of a constriction in the near field of the vocal folds: Physical modeling and experimental validation

Lucie Bailly
Département Parole et Cognition, GIPSA–lab, UMR5216 CNRS, INPG, UJF, Université Stendhal, 961 Rue de la Houille Blanche, Domaine Universitaire, BP 46, 38402 Saint Martin d'Hères Cedex, Laboratoire d'Acoustique de l' Université du Maine, UMR6613 CNRS, Université du Maine, Avenue Olivier Messiaen, 72085 Le Mans Cedex 09, France

Xavier Pelorson, Nathalie Henrich, and Nicolas Ruty
Département Parole et Cognition, GIPSA-lab, UMR5216 CNRS, INPG, UJF, Université Stendhal, 961 Rue de la Houille Blanche, Domaine Universitaire, BP 46, 38402 Saint Martin d'Hères Cedex, France

(Received 21 November 2007; revised 29 July 2008; accepted 30 July 2008)

The involvement of the ventricular folds is often observed in human phonation and, in particular, in pathological and or some throat-singing phonation. This study aims to explore and model the possible aerodynamic interaction between the ventricular and vocal folds using suitable *in vitro* setups allowing steady and unsteady flow conditions. The two experimental setups consist of a rigid and a self-oscillating vocal-fold replica, coupled to a downstream rigid ventricular-fold replica in both cases. A theoretical flow modeling is proposed to quantify the aerodynamic impact of the ventricular folds on the pressure distribution and thereby on the vocal-fold vibrations. The mechanical behavior of the vocal folds is simulated by a distributed model accounting for this impact. The influence of the ventricular constriction is measured in both flow conditions and compared to the model outcome. This study objectively evaluates the additional pressure drop implied by the presence of a ventricular constriction in the larynx. It is demonstrated that such constriction can either facilitate or impede the glottal vibrations depending on the laryngeal geometrical configuration. The relevance of using static or dynamic vocal-fold replicas is discussed. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977740]

PACS number(s): 43.75.Rs, 43.70.Bk, 43.70.Jt [AL] Pages: 3296–3308

I. INTRODUCTION

The ventricular folds, also called "false vocal folds", are two laryngeal structures located above the glottis and the laryngeal ventricle. They are not commonly involved during ordinary speech. Yet, their adduction has been observed in voice disorders such as ventricular dysphonia cases, where their interference in phonation may occur as a compensatory function (Nasri *et al.*, 1996; Pinho *et al.*, 1999). Although their physical properties (high viscosity and low stiffness) are different from those of biomechanical oscillators such as the vocal folds (Agarwal, 2004), their vibration has been observed during pathological (Nasri *et al.*, 1996) and some throat-singing productions found in Asian cultures—Mongolian *Kargyraa*, Tibetan *Dzo-ke* chants (Fuks *et al.*, 1998; Lindestad *et al.*, 2001; Sakakibara *et al.*, 2001, 2004)—and Mediterranean traditional polyphonies—Sardinian Cantu *A Tenore* (Henrich *et al.*, 2006).

The effects of the ventricular folds in phonation are still poorly understood. *In vivo* laryngeal examinations support the idea of a strong physical interaction between the ventricular and vocal folds (Fuks *et al.*, 1998; Lindestad *et al.*, 2001; Sakakibara *et al.*, 2001; Bailly *et al.*, 2007). Yet the physical nature of this interaction is unknown: is it related to an acoustical feedback, a mechanical conduction through the laryngeal mucosa, an aerodynamic drive,…? Few papers in the literature explore several hypotheses on the possible acoustic, mechanical, and/or aerodynamic effects underlying this interaction.

From an acoustical point of view, Fuks *et al.* (1998), Sakakibara *et al.* (2001, 2004), and Imagawa *et al.* (2003) studied the effect of including ventricular folds in the laryngeal tract using a two-mass model of the vocal folds. They showed how a supraglottal constriction contributes to regulate the voice quality specific to throat-singing techniques. Sakakibara *et al.* (2001) showed an increase in the acoustical power in the range of 1–3 kHz in an inverse filtered spectrum of *Kargyraa* voices. Imagawa *et al.* (2003) evaluated the acoustical consequences of change in the area of the supraglottal structures. They showed that an increase in the laryngeal ventricle area moves the third and fourth formants (F_3 and F_4) toward the second formant F_2 and sharpens it. A recent computational study suggested that the impingement of the glottal jet on the ventricular folds acts as a dipole source of sound adding high frequency components around 2–2.5 kHz, which are around the acoustic resonance frequency of the ventricle (Zhang *et al.*, 2002). Hirschberg *et al.* (1996) also discussed this dipole sound generation, implying an enhancement of the radiation of the chaotic sound. The ventricular folds could induce a stable feedback oscillation of the glottal jet, which might result into whistling. The hypothesis of a mechanical transmission consistent with Rayleigh's theory of surface waves has been discussed by Tsai *et al.* (2004) in the case of the *Sygyt* phonation. The aerodynamical aspects of the vocal-ventricular interaction and, in

 © 2008 Acoustical Society of America

particular, the impact of the ventricular folds on the laryngeal airflow have barely been investigated. Fuks *et al.* (1998) speculated that the driving of the ventricular folds could be a negative pressure generated by the airstream contained in the glottal pulse. They also measured higher oesophageal pressures when phonation switches from modal phonation to the vocal-ventricular mode. Yet, the following questions remain unanswered: What is the aerodynamic contribution of a downstream constriction near the vocal folds? How does such constriction affect glottal flow and pressure distribution inside the larynx? What does it induce on the vocal-fold self-oscillations?

Previous experimental investigations dealing with *in vitro* setups have provided an initial insight into the influence of the ventricular folds on the glottal airflow (Shadle *et al.*, 1991; Pelorson *et al.*, 1995; Agarwal, 2004; Bailly *et al.*, 2006; Kucinschi *et al.*, 2006). Using a rigid nonoscillating replica combining vocal folds and ventricular folds, Agarwal (2004) observed an increase or a decrease in the translaryngeal airflow resistance depending on the laryngeal geometry. The glottal jet curvature due to the Coanda effect typically observed through *in vitro* static replicas (Pelorson *et al.*, 1994, 1995; Hirschberg *et al.*, 1996; Hofmans, 1998; Hofmans *et al.*, 2003; Scherer *et al.*, 2001; Shinwari *et al.*, 2003; Erath and Plesniak, 2006; Kucinschi *et al.*, 2006) decreases with the presence of the ventricular folds (Shadle *et al.*, 1991; Kucinschi *et al.*, 2006). Flow visualization by Kucinschi *et al.* (2006) showed a downstream shift of the glottal separation point in the presence of the ventricular folds, inducing a conservation of the flow laminar properties over a longer distance. A significant pressure recovery associated with a reattachment of the jet flow is observed in Pelorson *et al.* (1995), in line with Miller *et al.* (1988). Therefore, to date, most studies have been conducted on *in vitro* rigid motionless replicas of the larynx from a (quasi-) steady point of view (Pelorson *et al.*, 1995; Agarwal, 2004; Kucinschi *et al.*, 2006).

In this paper, we present a theoretical and experimental work aiming at analyzing the fluid phenomena that may be induced by the presence of a constriction downstream and close to the vocal folds, similar to the ventricular folds. The parameter investigated is the aperture of the ventricular folds. First, simplified theoretical flow aspects accounting for this parameter are proposed in Sec. II. Then, their relevance is tested on suitable *in vitro* experimental devices described in Sec. III. Two different setups are designed to validate the theoretical outcome. The first one consists of two rigid and static vocal-fold and ventricular-fold replicas aligned; the second one relies on a deformable self-oscillating vocal-fold replica coupled to a downstream rigid and static ventricular-fold replica of adjustable geometry. Within the scope of this study, the ventricular folds are thus explored under nonoscillatory conditions. Yet, we expect the results to provide relevant clues to study the physical contingency of ventricular-fold self-sustained vibrations. Those replicas are not intended to reproduce a physiological reality but rather to test the proposed theoretical predictions. Finally in Sec. IV, theoretical flow predictions are compared to *in vitro* pressure measurements. The impact on the glottal vibrations is quantitatively evaluated on experimental data, and these results are compared with simulations using a two-mass model of the vocal folds accounting for the aerodynamic influence described in Sec. II.

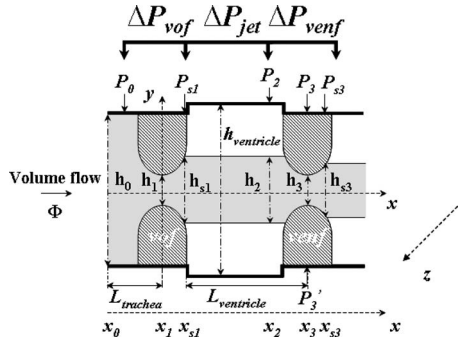


FIG. 1. Geometrical view of the laryngeal tube and relevant quantities of the aerodynamic study (vof: vocal folds; venf: ventricular folds).

II. FLOW THEORY

A geometrical representation of a simplified and schematic laryngeal tube is given in Fig. 1 along with all relevant parameters to model the aerodynamic jet behavior. The larynx is assumed to be symmetric with respect to the x and z axes. In the following, indices i correspond to specific positions along the x axis, as indicated in the figure. $h_i = h_i(x_i,t)$ and $A_i = W_i h_i$ respectively refer to the height and the cross-sectional area of the channel flow at the position x_i, $W_i = W(x_i)$ being the depth of the section along the z direction. h_{vof} (respectively h_{venf}, $h_{\text{ventricle}}$) corresponds to the aperture of the vocal folds (respectively the ventricular folds and the ventricle). Note that in this study, h_{vof} always equals h_1, whereas h_{venf} may differ from h_3. h_3^0 stands for a reference theoretical width of the flow in a no-ventricular-fold configuration. $P_i = P_i(x_i,t)$ represents the relative pressure predicted at x_i, as compared to the ambient atmospheric pressure.

The pressure distribution along the laryngeal tube is considered under the approximations of a laminar one-dimensional incompressible flow, separating at x_{s1} and x_{s3}. Three coupled subsystems are considered:

- the pressure drop across the vocal folds, $\Delta P_{\text{vof}} = P_0 - P_{s1}$,
- the emerging jet evolving in the ventricle, with a dissipation of kinetic energy $\Delta P_{\text{jet}} = P_{s1} - P_2$, and
- the pressure drop across the ventricular folds, $\Delta P_{\text{venf}} = P_2 - P_{s3}$.

The pressure P_{s3} is neglected, assuming a complete dissipation by turbulence downstream to the ventricular folds. Under these assumptions, the overall pressure drop ΔP governing the aerodynamics from the glottal inlet at x_0 to the flow separation at x_{s3} becomes

$$\Delta P = \Delta P_{\text{vof}} + \Delta P_{\text{jet}} + \Delta P_{\text{venf}} = P_0. \tag{1}$$

Equation (1) illustrates that the presence of a constriction in the flow channel can generate an additional pressure drop, $\Delta P_{\text{jet}} + \Delta P_{\text{venf}} = P_{s1}$. The differences ΔP_{jet} and ΔP_{venf} are commonly set to zero in the case of no additional constriction (Ruty *et al.*, 2007). If significant, this additional pressure

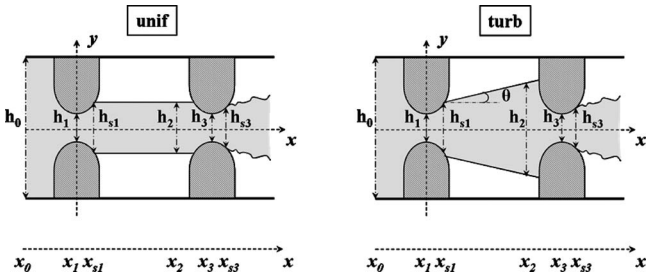


FIG. 2. Illustration of the hypotheses about the glottal jet expansion: (from left to right) uniform jet (unif); turbulent jet with $\delta \sim x$ and $\theta = 4°$ (turb), δ referred to as the glottal jet width.

drop is likely to alter the volume flow velocity, Φ, and produce a pressure recovery downstream the glottis. The theories detailed further below are an attempt to predict this pressure recovery using P_0 and h_1 as inputs to estimate consecutively A_{s1}, A_{s3}, A_2, Φ, P_{s1}, P_2, P_3, and $P_3{}'$ (see Fig. 1). They combine a separation model predicting A_{s1} and A_{s3}, a jet-flow geometrical model predicting A_2, and an aerodynamic model predicting Φ, P_{s1}, P_2, P_3, and $P_3{}'$.

A. Calculation of ΔP_{vof}

The estimation of the pressure drop on the inlet of the glottis $\Delta P_{vof} = P_0 - P_{s1}$ derives from the Euler equations governing the dynamics of inviscid flows coupled to a separation model. The expansion of the vocal folds to its downstream end causes the flow to separate from the wall at x_{s1}. For the sake of simplicity, a semiempirical model based on Liljencrants' *ad hoc* criterion is used as an alternative to a boundary-layer separation theory: $A_{s1} = 1.2 \times A_1$ (Ruty *et al.*, 2007). Neglecting the effects of wall vibrations and the volume airflow variations in time, the volume flow velocity Φ is constant, and it can be deduced from the overall pressure loss ΔP. Combining these assumptions with steady Bernoulli's equation applied between x_0 and x_{s1} yields to

$$\Delta P_{vof} = P_0 - P_{s1} = \frac{1}{2}\rho\Phi^2\left(\frac{1}{A_{s1}^2} - \frac{1}{A_0^2}\right), \tag{2}$$

where ρ is the air density.

B. Calculation of ΔP_{jet}

Two different jet flows are considered in the larynx: the jet downstream to the glottis and the jet downstream to the ventricular folds. Past the ventricular folds, the kinetic energy of the separated flow is assumed to be fully dissipated by turbulence (no reattachment), which theoretically comes to $P_{s3} = 0$. Understanding how the jet evolves from glottal separation to the ventricular inlet is crucial in order to predict accurately the pressure distribution along the channel.

Two hypotheses are formulated regarding the geometrical expansion of the glottal jet. Each of them is illustrated in Fig. 2.

- Hypothesis "unif" assumes a "uniform jet" without any expansion:

$$A_2 = h_2 W_2 = h_{s1} W_2. \tag{3}$$

- Hypothesis "turb," the so-called "turbulent jet" hypothesis, derives from a self-similar assumption of a free turbulent plane jet. A dimensional analysis leads to a linear growth of the jet width, δ, with increasing downstream distance regardless of the Reynolds number. Kundu (1990) reported from experimental data that a turbulent plane jet grows at a half-angle of 4°. A_2 calculation is then represented as follows:

$$A_2 = [h_{s1} + 2\tan 4\,^\circ\,(x_2 - x_{s1})]W_2. \tag{4}$$

Whatever the jet-flow expansion theory adopted, the dissipation ΔP_{jet} due to flow enlargement is either neglected (Hyp.a) or modeled by a quasisteady-pressure recovery determined from Eq. (5) (Hyp.b), in line with the estimations based on momentum considerations and the equation for diffuser flow suggested in Ishizaka and Flanagan (1972). This equation derives from the macroscopic mass and momentum conservation laws applied to the flow-detachment section A_{s1} and to section A_2 under the assumption that the pressure on section $A_2 - A_{s1}$ equals P_{s1} according to experimental observations (Ishizaka and Matsudaira, 1972; Blevins, 1992),

$$\Delta P_{jet} = -\frac{\rho}{2}\frac{\Phi^2}{A_{s1}^2}\left[2\frac{A_{s1}}{A_2}\left(1 - \frac{A_{s1}}{A_2}\right)\right]. \tag{5}$$

C. Calculation of ΔP_{venf}

The calculation of $\Delta P_{venf} = P_2$ is similar to the one of ΔP_{vof}. The divergent part of the ventricular folds decelerates the flow until separation at x_{s3}, determined by means of Liljencrants' semiempirical model. ΔP_{venf} is then estimated, thanks to the Bernoulli equation applied between x_2 and x_{s3}. The aerodynamical impact of the ventricular folds is only considered when the ventricular aperture h_{venf} drops below a critical height, h_3^0. This transition gap is estimated as the height of the channel flow, which would be theoretically expected at x_3 in the case without any ventricular constriction. If h_{venf} exceeds this aperture, it is then assumed that the presence of the ventricular folds does not alter the glottal jet, and therefore ΔP_{venf} is neglected,

$$\Delta P_{venf} = P_2 = \frac{1}{2}\rho\Phi^2\left(\frac{1}{A_{s3}^2} - \frac{1}{A_2^2}\right) \quad \text{if } h_{venf} < h_3^0,$$

$$\Delta P_{venf} = P_2 = 0 \quad \text{if } h_{venf} \geqslant h_3^0. \tag{6}$$

Three assumptions on flow are also considered. A potential quasisteady flow described by Eqs. (2) and (6) (Hyp.1), a viscous quasisteady flow (Hyp.2), and a potential unsteady flow (Hyp.3) are hypothesized. Equations (2) and (6) are modified to account for viscous flow effects (Hyp.2), whose contribution is no longer negligible for small constriction heights. The extra pressure loss induced by fluid viscosity is expressed by a Poiseuille correction reading

$$12\mu\int_{x_0}^{x_{s1}}\frac{\Phi(t)}{A^3(x,t)}dx$$

in Eq. (2) and

TABLE I. Summary of the different flow assumptions considered.

		Jet-flow geometrical expansion	
$\Delta P_{jet \neq 0}$ / $\Delta P_{jet = 0}$		"Uniform"	"Turbulent"
Flow dynamics	Quasi-steady Bernoulli + Flow separation	Hyp.1a$_{unif}$ / Hyp.1b$_{unif}$	Hyp.1a$_{turb}$ / Hyp.1b$_{turb}$
	Quasi-steady Bernoulli + Poiseuille viscous losses + Flow separation	Hyp.2a$_{unif}$ / Hyp.2b$_{unif}$	Hyp.2a$_{turb}$ / Hyp.2b$_{turb}$
	Unsteady Bernoulli + Flow separation	Hyp.3a$_{unif}$ / Hyp.3b$_{unif}$	Hyp.3a$_{turb}$ / Hyp.3b$_{turb}$

$$12\mu\int_{x_2}^{x_{s3}}\frac{\Phi(t)}{A^3(x,t)}dx$$

in Eq. (6), where μ is the air dynamic viscosity. Under Hyp.3, ΔP_{vof} and ΔP_{venf} are equal to the Bernoulli pressure corrected for flow unsteadiness due to air inertia. The additional term in the estimation of ΔP_{vof} and ΔP_{venf} reads

$$\rho\int_{x_0}^{x_{s1}}\frac{\partial}{\partial t}\frac{\Phi(t)}{A(x,t)}dx$$

and

$$\rho\int_{x_2}^{x_{s3}}\frac{\partial}{\partial t}\frac{\Phi(t)}{A(x,t)}dx,$$

respectively. Table I summarizes the different theories introduced.

Combining ΔP_{vof}, ΔP_{jet}, and ΔP_{venf}, one can predict the instantaneous pressure within the larynx. Under (Hyp.1a$_{unif}$) assumption, for instance, the volume airflow can be analytically calculated and the system {(2) and (6)} provides the following predictions for P_2 and P_3:

$$\frac{P_2}{P_0}=\frac{(A_{s1}^2-A_{s3}^2)}{(A_0^2-A_{s3}^2)}\frac{A_0^2}{A_{s1}^2}\quad \text{if } h_{venf}<h_3^0,$$

$$\frac{P_2}{P_0}=0\quad \text{if } h_{venf}\geqslant h_3^0\quad (\text{with } h_3^0=h_{s1}), \tag{7}$$

$$\frac{P_3}{P_2}=-0.44\frac{A_1^2}{A_1^2-A_3^2}\quad \text{if } h_{venf}<h_3^0,$$

$$\frac{P_3}{P_2}\quad \text{not defined if}\quad h_{venf}\geqslant h_3^0. \tag{8}$$

The turbulent jet assumption (Hyp.1a$_{turb}$) allows us to express the pressure recovery by geometrical ratios as well,

$$\frac{P_2}{P_0}=\frac{1}{1+\dfrac{A_{s3}^2A_2^2(A_0^2-A_{s1}^2)}{A_0^2A_{s1}^2(A_2^2-A_{s3}^2)}}\quad \text{if } h_{venf}<h_3^0,$$

$$\frac{P_2}{P_0}=0\quad \text{if } h_{venf}\geqslant h_3^0. \tag{9}$$

Alternatively, Agarwal (2004) proposed the use of another parameter, R, defined as

$$R=\frac{P_0-P_{s3}}{\Phi_{venf}}\frac{\Phi_{vof}}{P_0-P_{s3}}=\frac{\Phi_{vof}}{\Phi_{venf}}, \tag{10}$$

where Φ_{vof} refers to the base volume airflow calculated without ventricular folds in the larynx and Φ_{venf} refers to the volume airflow estimated in the presence of ventricular folds. Figure 3 presents two theoretical predictions of R as a function of h_{venf}/h_{vof}, deriving from the assumptions previously mentioned (Hyp.1a$_{turb}$ and Hyp.1a$_{unif}$). In each case, Eq. (10) leads to

$$R=\sqrt{\frac{\dfrac{1}{A_{s3}^2}-\dfrac{1}{A_0^2}+\dfrac{1}{A_{s1}^2}-\dfrac{1}{A_2^2}}{\dfrac{1}{A_{s1}^2}-\dfrac{1}{A_0^2}}}\quad \text{if } h_{venf}<h_3^0,$$

$$R=1\quad \text{if } h_{venf}\geqslant h_3^0. \tag{11}$$

Figure 3 presents the nonlinear relationship between R and h_{venf}/h_{vof} for $L_{ventricle}=26$ mm, $h_{vof}=2.90$ mm, and P_0

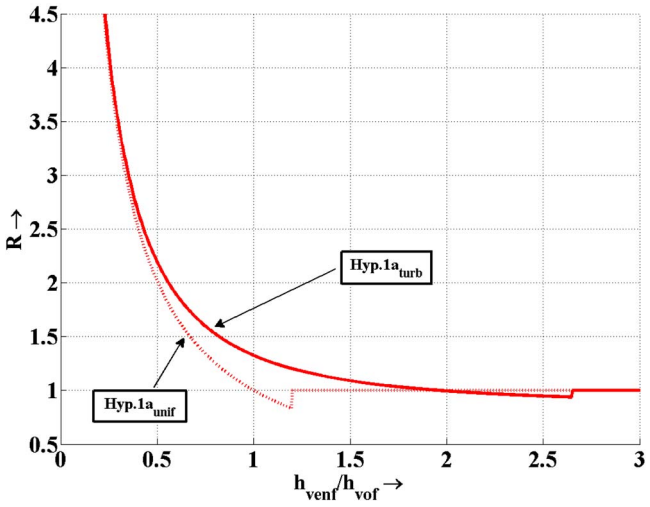


FIG. 3. (Color online) Theoretical predictions of the translaryngeal airflow resistance $R=\Phi_{vof}/\Phi_{venf}$ against h_{venf}/h_{vof} under two flow assumptions: Hyp.$1a_{unif}$ (dotted line) and Hyp.$1a_{turb}$ (solid line).

=500 Pa. It is interesting to note that these predictions qualitatively explain the experimental data presented by Agarwal (2004). More precisely R appears as a decreasing function of h_{venf}/h_{vof} and tends toward an asymptotic value of 1. Moreover, for a small region of h_{venf}/h_{vof}, R can become smaller than unity. This result is also in line with Kucinschi *et al.* (2006), who showed by means of flow visualization that the presence of the ventricular folds can accelerate and straighten the glottal jet [as previously observed by Shadle *et al.* (1991)], thus creating less energy dissipation.

III. MATERIAL AND METHOD

As an attempt to validate the above mentioned theoretical expectations, different experimental setups are designed, allowing steady or unsteady *in vitro* flow conditions.

A. Description of the laryngeal geometry

1. Anatomical dimensions

Tomographic anatomical observations approximate the gap separating the ventricular folds h_{venf} ranging from 2.3 to 8.3 mm (mean value of 6.1 mm) for male subjects phonating an /a/ in laryngeal mechanisms M_1 (equivalent to "modal" laryngeal register) and M_2 (equivalent to a "falsetto" laryngeal register) (Henrich, 2006; Roubeau *et al.*, 2007) and ranging from 4.0 to 6.8 mm (mean value of 5.5 mm) for male singers performing an /a/ (Hollien and Colton, 1969; Wilson, 1976; Agarwal, 2004; Agarwal *et al.*, 2003). The average distance from the superior-medial surfaces of the vocal folds to the minimum constriction height of the ventricular folds, $L_{ventricle}$ (see Fig. 1), reportedly ranges from 3.7 to 7.5 mm (mean value of 5.3 mm) for speakers and ranges from 3.3 to 6.1 mm (mean value of 4.6 mm) for singers. The ventricle height $h_{ventricle}$ (see Fig. 1) is in the range of 9.6–23.0 mm (mean value of 16.0 mm) for male speakers and 10.8–15.6 mm (mean value of 13.9 mm) for male singers. Those physiological data are provided in Table II.

2. Description of the laryngeal replicas

The static experimental setup consists of a 3:1 upscaled simplified replica of the human larynx. The laryngeal replica consists first of a rigid 25-mm-thick vocal-fold replica having a rounded shape (radius of curvature of 6 mm). The vocal-fold replica is coupled to a rigid 25-mm-thick ventricular-fold replica having a rounded shape (radius of curvature of 5 mm). The coupling between the vocal and the ventricular folds defines a ventricle having a circular cross-section of a constant diameter ($h_{ventricle}$=26 mm). The distance $L_{trachea}$ separating the glottal constriction to the pressure reservoir is 258 mm. The glottal aperture h_{vof}, as well as the ventricular aperture h_{venf}, can be controlled parametrically with an accuracy of $\pm$0.05 mm.

Except when mentioned, $L_{ventricle}$ is chosen as 26 mm, which is the closest distance allowed by the setup.

TABLE II. Comparative table: physiological characteristics of a male adult voice in speech and singing phonation [typical data obtained from Hollien and Colton (1969), Wilson (1976), Hirano *et al.* (1983), Agarwal (2004), Agarwal *et al.* (2003), and Bailly *et al.* (2007)] against geometrical and dynamical characteristics of the laryngeal replica.

	In vivo data	*In vitro* data
Scale	1	3
Glottal length, d, mean value (along the x direction)	4 mm	9 mm, 12 mm[a]
Glottal gap, h_{vof}, range of values	0–1.02 mm	0–5.95 mm
Ventricular gap, h_{venf}, range of values	0–9.0 mm	0.04–26 mm
Ventricle length, $L_{ventricle}$, range of values	3.3–7.5 mm	26–181 mm
Ventricle height, $h_{ventricle}$ range of values	9.6–23.0 mm	26 mm
Aspect ratio, $h_{venf}/L_{ventricle}$, range of values	0–2.7	0–1
Alimentation pressure, P_0, range of values	300–1000 Pa	100–2000 Pa
Reynolds number, Re, range of values	1700–3300[b]	1000–4600[b]
Strouhal number, Sr, range of values	0.01–0.03[b]	0.02–0.1[b]

[a]9 mm for the dynamical laryngeal replica, 12 mm for the static laryngeal replica.
[b]Estimated from alimentation pressure, glottal gap, h_{vof}=1 mm (mean value), and frequency of oscillations, F_0=150 Hz (mean value).

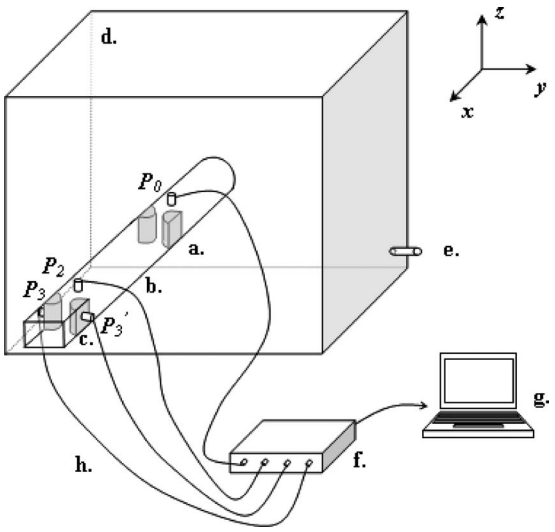


FIG. 4. Sketch of the flow facility and experimental setup using the static laryngeal replica (not to scale). (a) Static replica of the vocal folds, (b) cylindrical replica of the laryngeal ventricle, (c) static replica of the ventricular folds, (d) pressure reservoir (lungs), (e) air admission, (f) signal conditioner, (g) PC, and (h) pressure sensors.

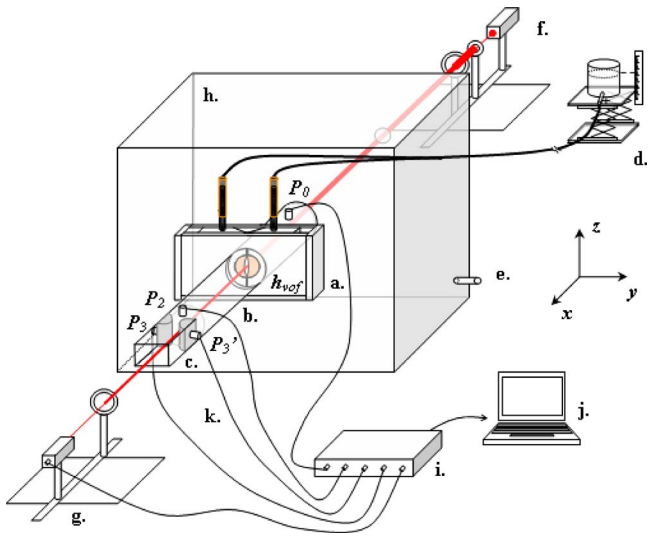


FIG. 5. (Color online) Sketch of the flow facility, the optical device, and the experimental setup using the dynamical laryngeal replica (not to scale). (a) Dynamical replica of the vocal folds, (b) cylindrical replica of the laryngeal ventricle, (c) static replica of the ventricular folds, (d) water pressure reservoir, (e) air admission, [(f) and (g)] laser diode, optical lenses, and photo sensor, (h) pressure reservoir (lungs), (i) signal conditioner, (j) PC, and (k) pressure sensors.

In order to realize unsteady flow conditions, another experimental setup has been used. It consists of a 20-mm-thick replica of the vocal folds made of rubber filled with water under pressure. For a sufficient alimentation pressure P_0, the vocal-fold replica can be driven into self-sustained oscillations. Alimentation pressures, vibration amplitude, and oscillation frequencies are comparable with human phonation. Details of this setup can be found in Ruty *et al.* (2006).

The pressure of the water inside the self-oscillating replica is chosen as 5000 Pa. This corresponds to an optimal oscillatory behavior as observed in Ruty *et al.* (2007). Under this configuration, the glottal longitudinal length, d, is 9 mm in the dynamical replica.

The ventricular folds and ventricle replicas are identical to those used in the static configuration. In order to avoid strong acoustical coupling with the downstream elements (Zhang *et al.*, 2006), the inlet length L_{trachea} is reduced to 119 mm. This induces a first quarter-wave acoustical resonance at about 714 Hz, which is much higher than the fundamental frequency of the airflow oscillations that does not exceed 200 Hz.

The correspondence between *in vivo* data and *in vitro* data allowed by our device is detailed in Table II. Dynamic similarity is preserved by relevant dimensionless Reynolds ($\text{Re}=U_1 h_{\text{vof}}/\upsilon$) and Strouhal ($\text{Sr}=F_0 d/U_1$) numbers, where the characteristic mean flow velocity U_1 is approximated by $U_1=\sqrt{2\Delta P_{\text{vof}}/\rho}$, υ is the air kinematic viscosity, and F_0 is the mean oscillation fundamental frequency.

B. Experimental procedure on the static laryngeal replica

The experimental setup is illustrated in Fig. 4. A pressure reservoir supplies the setup. The dynamic pressures upstream, P_0, within the jet, P_2, and within the ventricular-fold replica, P_3, are measured using Kulite (XCS-0.93-0.35-Bar-G) or Endevco (8507C) pressure sensors supplied by a Labor-Netzgerat power supply (EA-3005S).

The pressure sensor measuring P_0 is located 24 mm upstream of the vocal-fold replica. The sensor measuring P_2 is set 3 mm upstream of the ventricular-fold replica. P_3 is measured at the minimum constriction position within the ventricular-fold replica. An additional pressure sensor P_3' was also added to check for an asymmetry of the flow.

All pressure sensors are calibrated against an electronical manometer (Aschcroft XLdp) with typical accuracy of ± 5 Pa. Electrical signals are postprocessed using a preamplifier/conditioning board (National Instruments SXCI-1121) connected to a PC through a National Instruments BNC-2080 card and a National Instruments PCI-MIO-16XE acquisition card. The acquired data are processed using the Labview7 software (National Instruments). The sampling frequency for all acquisitions is 10 kHz. The pressure measurements are repeated several times for each geometrical configuration. The pressure signal is smoothed using a low-pass Butterworth filter (20 Hz cutoff frequency).

The variable apertures of both the vocal-fold and the ventricular-fold replica were measured using calibrated plates with an accuracy of 0.05 mm. P_0 is controlled and maintained constant thanks to the manometer mentioned above.

C. Experimental procedure on the dynamical laryngeal replica

1. Pressure and geometrical dynamics measurement

The unsteady dynamical experimental setup is illustrated in Fig. 5. The pressure sensor measuring P_0 is located 35 mm upstream of the vocal-fold replica, whereas the other pressure sensors were located as previously.

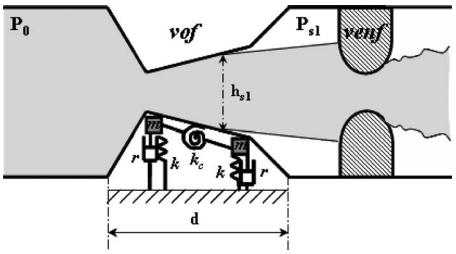


FIG. 6. Sketch of the two-mass model of the vocal folds combined to the theoretical flow description of the ventricular-fold influence under Hyp. $1a_{turb}$.

During an experimental round, the upstream pressure P_0 is continuously increased until self-sustained oscillations of the vocal-fold replica appear. An optical system consisting of a laser diode (P. Fontaine dc amplifier FTN2515 supply) aligned with a photosensor (BPW34, Solartron dc power supply) allows us to quantify the vocal-fold replica opening during oscillation. The optical system does not measure the opening area but the minimum glottal aperture at the mid-width of the vocal-fold replica along the flow direction, h_{vof}. Details of this setup can be found in Ruty *et al.* (2007). Typical accuracy for the measure of h_{vof} is ± 0.01 mm.

The mechanical response of the deformable replica is obtained using a pressure driver unit (ERSTU-100) added in the vicinity of the vocal-fold replica. It provides a sinusoidal excitation whose frequency varies from 45 to 500 Hz in steps of 1 Hz. The opening h_{vof} for each acoustical excitation is captured by means of the optical device previously detailed. Mechanical resonance frequencies, quality factors, and equilibrium positions are then extracted (Ruty *et al.*, 2007). From these parameters mechanical constants for the theoretical models can be easily inferred.

The experimental phonation onset pressure P_{onset} represents the minimum upstream pressure necessary to sustain the vocal-fold replica oscillations. The upstream pressure for which oscillations disappear in a decreasing-flow sweep corresponds to the phonation offset pressure P_{offset}. Both pressures are measured from the recorded pressure signal P_0, modulated by the oscillation, using a periodicity detection method. The fundamental frequency of vocal-fold replica oscillation is also extracted from the measurements via a 0.05 s windowing. The periodicity detection method is based on autocorrelation and peak detection techniques classically used in voice signal processing (Henrich *et al.*, 2004).

2. Simulation of the vocal-fold dynamics using a two-mass model

In order to analyze the experimental data with the self-oscillating vocal-fold replica, a theoretical model for the self-sustained oscillations is needed. For the sake of simplicity, a reduced "two-mass" mechanical model is used (Lous *et al.*, 1998), as shown in Fig. 6. It is controlled by a set of mechanical input parameters: the mass (m), stiffness (k,k_c), and damping (r). These parameters are directly deduced from the mechanical response of the deformable replica.

The aerodynamical contribution of the ventricular folds downstream of the glottis is accounted for by considering the pressure past the flow separation point that does not equal the atmospheric pressure, i.e., $P_{s1} \neq 0$. P_{s1} is predicted using the theoretical expectation detailed in Sec. II. The onset subglottal pressure and the fundamental frequency of oscillation are predicted by applying a linear stability analysis to the mechanical model (Van Hirtum *et al.*, 2007; Ruty *et al.*, 2007).

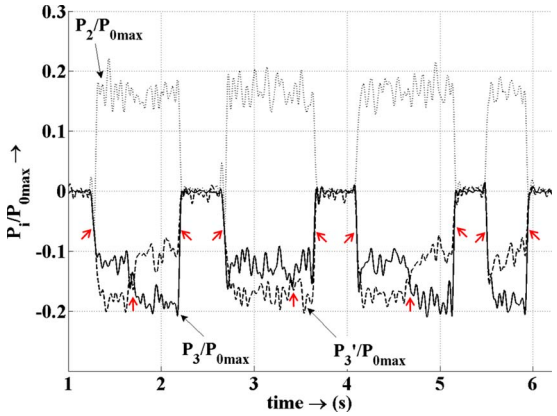


FIG. 7. (Color online) Experimental normalized pressures $P_2/P_{0\,max}$ (dotted line), $P_3/P_{0\,max}$ (solid line), and $P_3'/P_{0\,max}$ (dashed line) as functions of time acquired on the static laryngeal replica. Four repetitions of a measurement with $L_{ventricle}=31.5$ mm, $h_{vof}=3.25$ mm, and $h_{venf}=5.95$ mm, i.e., $h_{venf}/h_{vof}\sim 1.8$ and $P_{0\,max}=109$ Pa. In the case of the first and third repetitions, note the changing asymmetry of the flow at the level of the ventricular-fold constriction. Small arrows indicate time sequences for which $P_3/P_{0\,max}$ and $P_3'/P_{0\,max}$ are close.

IV. RESULTS AND DISCUSSION

In this section, we will compare the predictions of the theoretical models and the experimental data. First, the influence of the ventricular folds on the glottal flow and on the pressure distribution in the larynx is considered. A second important feature is the phonation threshold pressures and the fundamental frequency. In principle, up to 12 theoretical expectations have been tested systematically. However, for the sake of clarity, only a few typical illustrative predictions will be presented.

A. Static laryngeal replica

1. Experimental observations

Figure 7 presents a typical example of the experimental data acquired on the static laryngeal replica. The parameters used are $L_{ventricle}=31.5$ mm and $h_{venf}/h_{vof}\sim 1.8$. In this example, the upstream alimentation pressure P_0 is increased from 0 Pa to a permanent flow threshold maintained at around 100 Pa. The measurement is repeated four times under the same experimental conditions. P_2, P_3, and P_3' are normalized by the average threshold $P_{0\,max}$ and are displayed as functions of time. As preliminary observations, a pressure recovery is noticeably evidenced: P_2 increases to about 20% of the alimentation pressure P_0. The pressures measured for P_3 and P_3' are negative, which is likely to correspond to a Bernoulli effect. An interesting feature can be highlighted: P_3 and P_3' variations reveal a flow instability within the ventricular-fold replica for a stationary glottal flow (as indicated by stationary values of P_0 and P_2). These discrepancies

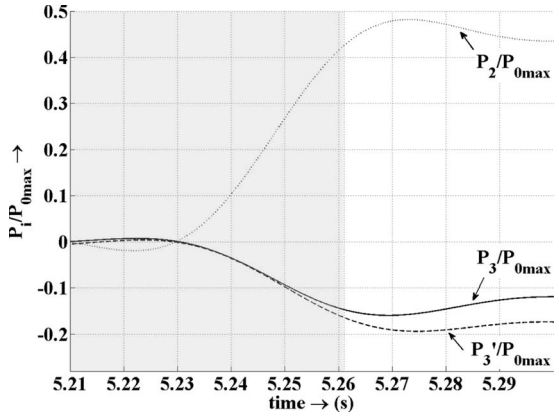


FIG. 8. Experimental normalized pressures $P_2/P_{0\,\max}$ (dotted line), $P_3/P_{0\,\max}$ (solid line), and $P_3'/P_{0\,\max}$ (dashed line) as functions of time acquired on the static laryngeal replica. $L_{\text{ventricle}}=31.5$ mm, $h_{\text{vof}}=3.25$ mm, and $h_{\text{venf}}=3.10$ mm, i.e., $h_{\text{venf}}/h_{\text{vof}}\sim 0.9$ and $P_{0\,\max}=296$ Pa.

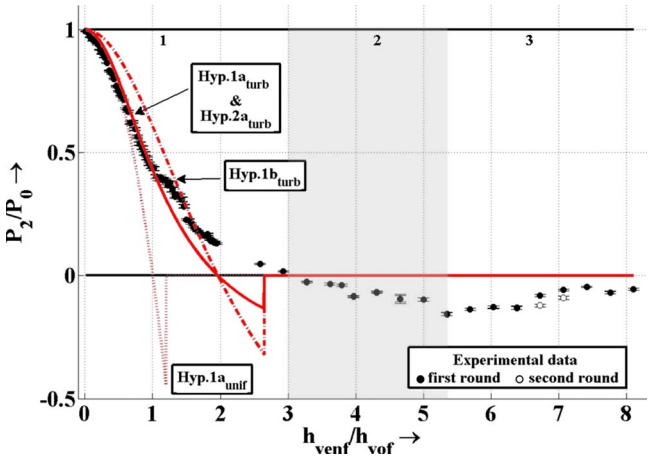


FIG. 9. (Color online) Normalized pressure recovery P_2/P_0 measured for two experimental rounds using the static laryngeal replica, displayed as functions of $h_{\text{venf}}/h_{\text{vof}}$ and plotted against theoretical predictions. $L_{\text{ventricle}}=26$ mm, $h_{\text{vof}}=2.90$ mm, and $P_{0\,\max}=500$ Pa. Error bars correspond to the standard deviation of the measured data.

can be understood as the result of a Coanda effect, which induces a skewing of the jet flow to one or the other side of the ventricular-fold replica (Pelorson *et al.*, 1995; Kucinschi *et al.*, 2006). As a matter of fact, the measured $P_3/P_{0\,\max}$ and $P_3'/P_{0\,\max}$ reveal not only an asymmetrical flow but also that the asymmetry can change from one side to the other. The dynamical aspect of the jet skewing is thus evidenced, consistently with Kucinschi *et al.* (2006), who also observed a jet downstream to the ventricular folds "bistable" in direction for a narrow geometrical configuration $h_{\text{venf}}/h_{\text{vof}}$.

Figure 8 displays time dependent pressure measurements performed for a smaller ventricular gap ($h_{\text{venf}}/h_{\text{vof}}\sim 0.9$). P_0 is increased from 0 Pa up to a steady-pressure value of 296 Pa. Experimental data exhibit a significant pressure recovery ($P_2\ \alpha\ 46\%$ of P_0), as well as a Coanda effect at the ventricular-fold replica. Note that for P_0 below around 273 Pa (see the gray area in Fig. 8), $P_3/P_{0\,\max}$ and $P_3'/P_{0\,\max}$ are identical with respect to the measurement accuracy (± 5 Pa), suggesting a time sequence where the glottal jet expands symmetrically before the Coanda phenomenon is established. This result is another dynamical feature of the Coanda effect. In particular, our results support the work by Pelorson *et al.* (1994), Hirschberg *et al.* (1996), Hofmans (1998), and Hofmans *et al.* (2003), who suggested that the Coanda effect is not important during a transient and suggested that a sufficiently unsteady flow might delay the onset of the Coanda effect.

2. Pressure drop at the vocal folds

In this section we focus on the perturbations induced on the pressure field through the static vocal-fold replica. An example of results for $L_{\text{ventricle}}=26$ mm and $h_{\text{vof}}=2.90$ mm is presented in Fig. 9. In this example $h_{\text{venf}}/h_{\text{vof}}$ is parametrically varied from 0.02 up to 1.95 by changing the ventricular-fold aperture h_{venf} by steps of 0.03. For larger values of $h_{\text{venf}}/h_{\text{vof}}$, a larger step, 0.3, is used. The alimentation pressure P_0 is maintained constant as 500 Pa. For each geometrical configuration $h_{\text{venf}}/h_{\text{vof}}$, pressure recordings have been repeated twice.

As shown in Fig. 9, the pressure P_2 decreases with $h_{\text{venf}}/h_{\text{vof}}$. More precisely, different trends can be distinguished with regard to the generated perturbation on the pressure distribution (areas labelled {1,2,3} in Fig. 9):

- For $h_{\text{venf}}/h_{\text{vof}}$ in the range from ~ 0 up to ~ 3 (area 1), a positive pressure P_2 is measured, reaching a P_0 value for very small values of $h_{\text{venf}}/h_{\text{vof}}$. As a consequence, the pressure drop at the vocal folds decreases down to zero when $h_{\text{venf}}/h_{\text{vof}}$ tends toward 0.
- For $h_{\text{venf}}/h_{\text{vof}}$ in the range from ~ 3 up to ~ 5 (area 2), the pressure P_2 becomes negative. In this range, P_2 still decreases with $h_{\text{venf}}/h_{\text{vof}}$ down to a minimum at $h_{\text{venf}}/h_{\text{vof}}\sim 5.3$. This result is in line with previous experimental observations but with different quantitative values [the minimum is reported at $h_{\text{venf}}/h_{\text{vof}}\sim 2$ for Agarwal (2004)].
- For $h_{\text{venf}}/h_{\text{vof}}$ higher than ~ 5 (area 3), the pressure P_2 is still negative but increases with $h_{\text{venf}}/h_{\text{vof}}$.

Two asymptotic limits $P_2/P_0=1$ (i.e., $\Delta P_{\text{vof}}+\Delta P_{\text{jet}}=0$) and $P_2/P_0=0$ (i.e., $\Delta P_{\text{vof}}+\Delta P_{\text{jet}}=P_0$) illustrate the physical extent of the pressure recovery qualitatively expected for very high and very low degrees of ventricular constriction, respectively. Four theoretical predictions are plotted in Fig. 9 against the experimental observations. Three of them consider a turbulent jet assumption (Hyp.$1a_{\text{turb}}$, Hyp.$2a_{\text{turb}}$, and Hyp.$1b_{\text{turb}}$), while the fourth relies on a uniform jet assumption (Hyp.$1a_{\text{unif}}$).

The P_2/P_0 decrease with increased $h_{\text{venf}}/h_{\text{vof}}$ is qualitatively well predicted by the four predictions. However, an abrupt transition to $P_2/P_0=0$ is predicted by all theories as soon as $h_{\text{venf}}\geqslant h_3^0$, as described in Eq. (6).

From a quantitative point of view, the theoretical prediction based on the (Hyp.$1a_{\text{unif}}$) assumption provides the worst estimation except for extreme values of $h_{\text{venf}}/h_{\text{vof}}$. Considering an empirical quasistationary pressure recovery in addition (Hyp.$1b_{\text{turb}}$), like the one suggested by Ishizaka and Flanagan (1972) although providing a better estimate on average, does obviously overestimate the pressure recovery past the glottis. For very small values of $h_{\text{venf}}/h_{\text{vof}}$, the estimation of P_2/P_0 is even worse than the uniform jet assump-

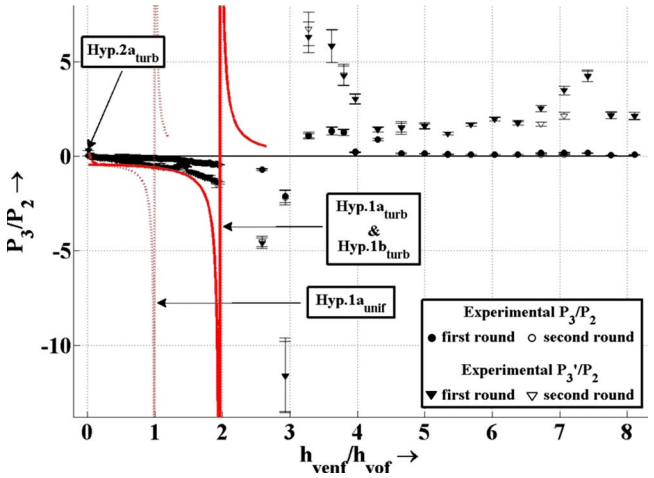


FIG. 10. (Color online) Normalized pressure ratios P_3/P_2 and P_3'/P_2 measured for two experimental rounds using the static laryngeal replica, displayed as functions of h_{venf}/h_{vof} and plotted against theoretical predictions. $L_{ventricle}$=26 mm, h_{vof}=2.90 mm, and $P_{0\,max}$=500 Pa. Error bars correspond to the standard deviation of the measured data.

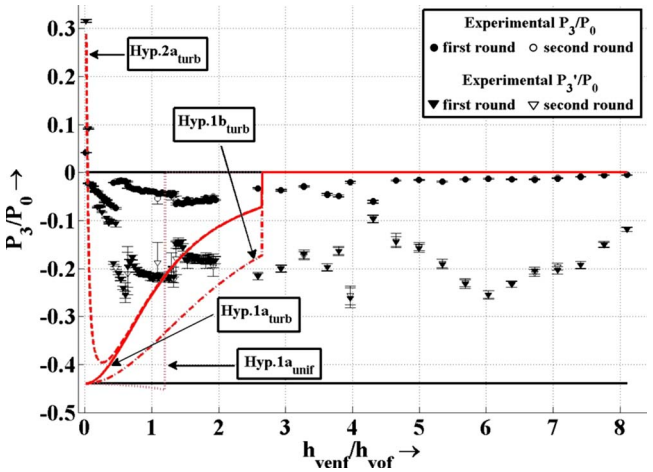


FIG. 12. (Color online) Normalized pressure ratios P_3/P_0 and P_3'/P_0 measured for two experimental rounds using the static laryngeal replica, displayed as functions of h_{venf}/h_{vof} and plotted against theoretical predictions. $L_{ventricle}$=26 mm, h_{vof}=2.90 mm, and $P_{0\,max}$=500 Pa. Error bars correspond to the standard deviation of the measured data.

tion (Hyp.1a$_{unif}$). Lastly, the theoretical prediction based on the turbulent jet assumption (Hyp.1a$_{turb}$) provides the best estimation for the experimental data. A relative error below 8% is observed for configurations such as $h_{venf}/h_{vof}<1$. Therefore, this turbulent jet geometrical assumption seems more suitable for this study. Taking viscous pressure losses (Hyp.2a$_{turb}$) into account has very little effect and does not improve the accuracy of the predictions.

3. Pressure drop at the ventricular folds

In addition to P_2/P_0, the pressure at the ventricular folds is quantified here using ratios P_3/P_2 and P_3'/P_2. These parameters are of importance as indicators of possible ventricular-fold self-sustained oscillations.

Figure 10 displays experimental ratios P_3/P_2 and P_3'/P_2 as functions of h_{venf}/h_{vof}. Two areas are exhibited:

- For h_{venf}/h_{vof} up to ~3, the experimental data show negative pressures measured for both P_3 and P_3', confirming a Bernoulli effect, except for two data points close to the ventricular-fold replica full adduction (h_{venf}/h_{vof}=0.02 and h_{venf}/h_{vof}=0.05), as enhanced in Fig. 11. It is also interesting to note in Fig. 11 that for very small geometrical configurations such as $h_{venf}/h_{vof}<0.5$, P_3/P_2 and P_3'/P_2 are close. Kucinschi *et al.* (2006) similarly showed that a narrow geometrical configuration tends to straighten out the jet downstream to the vocal folds. As h_{venf}/h_{vof} increases, the difference between P_3 and P_3' increases and even exceeds 100% for large values. Such a case is likely to be due to a fully established Coanda effect.
- For h_{venf}/h_{vof} exceeding 3, positive values of P_3/P_2 and P_3'/P_2 are measured. It must be noted that the scattering of measured values increases significantly, probably revealing a fully turbulent behavior of the flow.

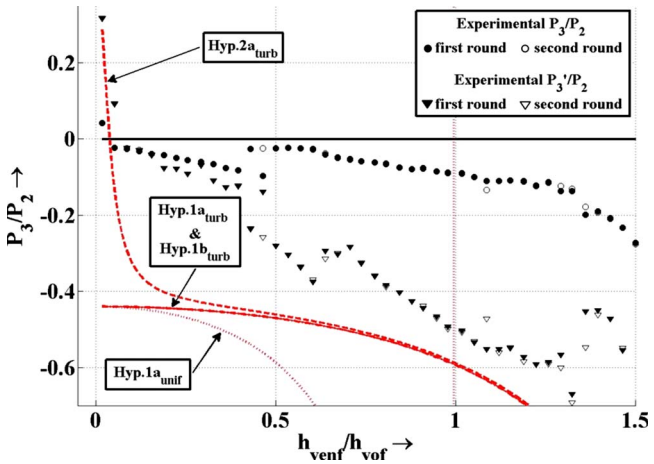


FIG. 11. (Color online) Close-up of Fig. 10.

Other illustrative parameters, P_3/P_0 and P_3'/P_0, as functions of h_{venf}/h_{vof} are displayed in Fig. 12. It is shown that little of the alimentation pressure is effectively recovered (15% on average). The four theoretical predictions are also presented on Figs. 10–12. Note that the prediction of P_3/P_2 is plotted until the configuration for which P_2=0. Qualitatively, the decrease in P_3/P_2, as well as the drastic change in sign past a critical value of h_{venf}/h_{vof}, is well predicted. Considering an additional pressure recovery (Hyp.2b$_{turb}$) due to the jet reattachment has very little influence on the estimation of P_3/P_2. As a matter of fact, the predictions based on assumptions Hyp.1a$_{turb}$ and Hyp.1b$_{turb}$ are identical. As shown on Figs. 11 and 12, taking into account viscosity effects in the flow (Hyp.2a$_{turb}$) yields a positive prediction of P_3 for extremely small values of h_{venf}/h_{vof}, in line with experimental observations.

Despite these similar qualitative trends, quantitative departures with the measurement are observed for all theories. These huge discrepancies are likely to be due to the presence of a fully turbulent flow behavior, which is not accounted for by any of the predictions presented here. As a matter of fact, contrarily to P_2, the sensor heads measuring P_3 and P_3' are located at a distance exceeding ten glottal diameters. Such a distance corresponds usually to a fully developed turbulent jet (e.g., Goldstein, 1976).

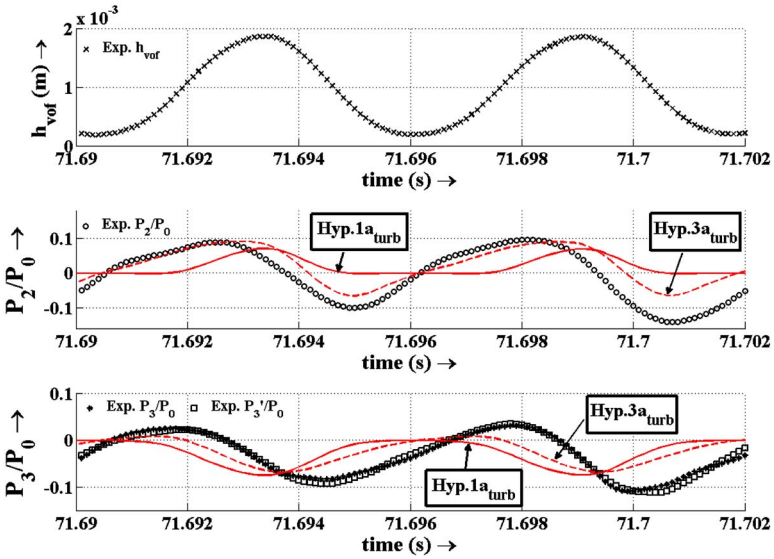


FIG. 13. (Color online) Measured data of h_{vof} and P_2, P_3, and P_3' normalized by the alimentation pressure P_0 as functions of time using the dynamical vocal-fold replica with respect to simulations with different jet assumptions. $L_{ventricle}$=26 mm, h_{venf}=3.50 mm, and h_{vof} =0.95 mm at rest and h_{vof}=0.97 mm on the average at the equilibrium position.

B. Dynamical laryngeal replica

1. Influence of the constriction on the pressure distribution

As an illustrative example, we present results obtained for h_{venf}=3.5 mm. The glottal replica aperture at rest was 0.95 mm. The constant alimentation pressure is chosen as P_0=1150 Pa. For such a configuration, the latex replica is subject to self-sustained oscillations at a frequency of 176 Hz. Typically, during an oscillation period the ratio h_{venf}/h_{vof} could vary between 1.8 and 20.

Figure 13 illustrates the pressure modulations generated by the vocal-fold self-sustained oscillations. Unsteady effects are clearly evidenced, particularly by the important phase difference between h_{vof} and P_2 and between P_2 and P_3. It is interesting to note that both P_3 and P_3' have very close values. Thus, during the dynamic experiments, the potential jet asymmetry seems to have much less effects than what was observed during the static experiments (see Sec. IV A 1). This result is in line with previous unsteady flow experiments (Pelorson *et al.*, 1994).

Consistent with the static experiments, the experimental data show that the measured pressures P_2, P_3, and P_3' are alternatively positive or negative. Negative values for P_3 and P_3' are likely to correspond to a Bernoulli effect on the ventricular-fold replica.

For the sake of clarity, only two theoretical predictions are presented here. The first one, based on the assumption (Hyp.$1a_{turb}$), is considered because of its performance during the static experiments. To this latter prediction, unsteady effects have been added, which correspond to assumption (Hyp.$3a_{turb}$). Note that the uniform jet assumption (Hyp.$1a_{unif}$) would, in such a configuration, always predict a zero pressure for both P_2 and P_3 since h_{venf} is always larger than the jet width h_{s1}.

Considering the prediction of P_2, the turbulent jet theory in quasisteady formulation (Hyp.$1a_{turb}$) provides a poor agreement with the measured data. Important amplitude discrepancies (of order of 60%) are observed as well as phase differences (of order of 60°). In comparison, the turbulent jet assumption, which accounts for unsteadiness (Hyp.$3a_{turb}$), provides a much better amplitude prediction (of order of 20%). Yet, although quite smaller, a phase lag is still observed (of order of 30°).

Considering the prediction of P_3, none of the theoretical predictions can explain quantitatively the measured data. Obviously, in this case, the turbulent viscous dissipation and unsteady phenomena within the jet are not correctly taken into account in the theories tested in this part. The amplitude error may be also related to the crude approximation of the glottal replica cross-sectional area, which has a more complex shape than a rectangular one.

2. Influence of the constriction on the vocal-fold self-oscillations

Lastly, we consider two parameters. Figure 14 presents the experimental oscillation threshold pressures $P_{threshold}$ (namely, P_{onset} and P_{offset}) and fundamental frequencies F_0 averaged on two series of measurements. h_{venf} ranges from 0.04 to 22 mm. h_{vof_eq} is defined as the equilibrium position,

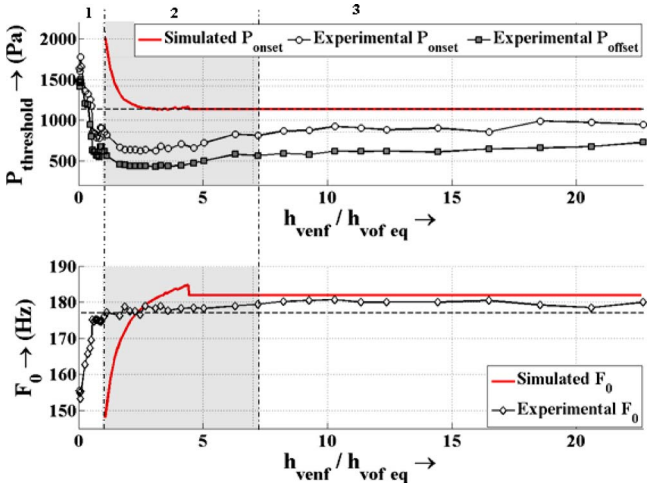


FIG. 14. (Color online) Measured threshold pressures and fundamental oscillation frequencies with corresponding two-mass model simulations, displayed as functions of h_{venf}/h_{vof_eq} for $L_{ventricle}$=26 mm. Experimental baselines P_{onset}=1136 Pa and F_0=177 Hz measured in a configuration without ventricular-fold replica attached are plotted in dashed lines.

namely, the mean value of h_{vof} during vocal-fold oscillations. h_{vof_eq} experimental value is averaged for all h_{venf} apertures allowing the optical detection of h_{vof} and equals 0.97 mm. For reference the onset pressure, offset pressure, and fundamental frequency measured without any ventricular-fold replica (i.e., for $h_{venf}=h_{ventricle}$) are, respectively, 1136 Pa, 833 Pa, and 177 Hz.

Some interesting trends can be observed. Firstly, for all the geometrical configurations explored here, experimental phonation offset pressures are found to be lower than experimental phonation onset pressures, which is in line with the hysteretic behavior noted by Lucero (1999) and Ruty *et al.* (2007).

Secondly, the experimental results concerning the influence of the geometrical configuration h_{venf}/h_{vof_eq} can be divided into three trends, referring to three different areas labeled in Fig. 14:

- For h_{venf}/h_{vof_eq} ranging up to 1 (area 1), the onset pressure is found greater or comparable to the threshold pressure obtained without the ventricular-fold replica attached. This implies that a severe constriction of the ventricular folds, relative to the glottal width, impedes the vocal-fold replica vibration. Note that this behavior co-occurs with an increase of the fundamental frequency with h_{venf}/h_{vof_eq}.
- For h_{venf}/h_{vof_eq} ranging between 1 and 7 (area 2), the onset pressure is found to be significantly lower than the one measured without ventricular-fold replica. These results suggest that within this geometrical range, the presence of ventricular-fold replica favors the vocal-fold replica oscillations. Such an effect has been reported by Agarwal (2004), suggesting that the ventricular folds may facilitate the phonation by increasing the airflow through the larynx. The fundamental frequency remains almost constant in this area.
- For h_{venf}/h_{vof_eq} higher than 7 (area 3), the onset pressure remains almost constant. The fundamental frequency is not altered by the geometrical configuration either.

The experimental results are compared to theoretical predictions using the vocal-fold two-mass model described in Sec. III C 2. The mechanical parameters of the simulation are obtained in order to fit the onset pressure and fundamental frequency measured without any ventricular folds attached. Namely, the parameters were

$$m = 1.90 \times 10^{-4}\ \text{kg},$$

$$k = 2.33 \times 10^{2}\ \text{N m}^{-1},$$

$$k_c = 1.16 \times 10^{2}\ \text{N m}^{-1},$$

$$r = 6.79 \times 10^{-3}\ \text{N s m}^{-1}.$$

All other parameters (alimentation pressure, initial glottal area, and geometry) were directly taken similar to those of the experiments. First, for all experimental cases the two-mass model predicted self-sustained oscillations except for $h_{venf}/h_{vof_eq}<1$ where no oscillations were possible. A qualitative decrease in the phonation onset pressures is predicted by the two-mass model, however, for a smaller range (i.e., h_{venf}/h_{vof_eq} between 1 and 4 instead of between 0.04 and 4 experimentally). For $h_{venf}/h_{vof_eq}>4$, the phonation onset pressures are predicted constant and equals the configuration without the ventricular folds. The fundamental-frequency increase with h_{venf}/h_{vof_eq} is quantitatively well simulated.

V. CONCLUSIONS

The aim of this both experimental and theoretical study was to evaluate the potential influence of a ventricular constriction during voicing. Several theoretical predictions based on different flow assumptions have been proposed and tested against an *in vitro* experimental setup under steady or unsteady flow conditions.

The conclusions that can be drawn from this study are as follows:

- Experimentally, very different results are obtained when considering either a static or a dynamical device for the vocal folds. This is observed as the occurrence not only of unsteady effects such as the phase shift between the glottal aperture and the pressure signals but also of the stability of the jet flow. Asymmetrical flows, such as the Coanda effect, which are typically observed for steady flow conditions (Pelorson *et al.*, 1994, 1995; Hirschberg *et al.*, 1996; Hofmans, 1998; Hofmans *et al.*, 2003; Scherer *et al.*, 2001; Shinwari *et al.*, 2003; Kucinschi *et al.*, 2006), seem quite limited under unsteady flow conditions created by a dynamical vocal-fold replica. While asymmetrical flow might indeed occur during phonation, for example, due to some asymmetries in the vocal-fold geometry, the relevance of the Coanda effect in phonation seems therefore doubtful.
- The geometrical configuration is shown to be a critical parameter, thereby confirming earlier studies (Agarwal, 2004; Kucinschi *et al.*, 2006). It is found that under a certain configuration, the presence of the ventricular folds can favor the oscillations of the vocal folds. On the other hand, with respect to the vocal-fold aperture, very narrow ventricular folds tend to perturbate or even suppress the vocal-fold vibrations. This is due to a significant pressure recovery that occurs past the glottal channel and which therefore reduces the transglottal pressure drop. Such geometrical configurations occur when the ventricular-fold aperture are smaller or equal to the glottal aperture. In such a case, a fundamental-frequency increase with h_{venf}/h_{vof} is observed. Although such a configuration is not expected in normal speech, it might be relevant during Asian vocal productions or Sardinian *A Tenore* singing where even a total closure of the ventricular folds is observed (Fuks *et al.*, 1998; Sakakibara *et al.*, 2004; Bailly *et al.*, 2007). Some pathological cases might also be concerned by such behavior. These trends have been observed experimentally. Two-mass model simulations combined to the theoretical flow description of the ventricular-fold influence demonstrate that these trends are mainly related to aerodynamic effects.
- Whether or not the ventricular folds can self-oscillate is still unclear. While the ventricular folds are *a priori* poor biomechanical vibrators (Agarwal, 2004), jet reattachment

and Bernoulli effect are indeed observed on the ventricular-fold replica. The pressure drop at the ventricular folds, which can be in first approximation estimated by P_2, can reach a considerable proportion of the alimentation pressure (e.g., 70% at $h_{venf}/h_{vof}=0.5$). This latter point has, of course, to be more extensively studied for a larger range of plausible laryngeal configurations. In particular, an *in vitro* experimental investigation including a deformable ventricular-fold replica, and thus capable of self-sustained oscillations, is currently under study.

- Theoretically, the prediction of all the above aerodynamical effects for the purpose of synthesis, for example, remains quite difficult. More precisely, it was shown that a simple Bernoulli theory based on the assumptions of an empirical flow separation, a linear growth of the glottal jet width, and unsteady flow effects provides a reasonable estimation of transglottal pressure drop. The glottal jet geometrical expansion appears then to be the most critical parameter. Due to the cascade of successive assumptions, the prediction of the pressure downstream remains only qualitative. At the level of the ventricular constriction, all theoretical predictions fail to explain the measured steady or unsteady data. Most probably, the high viscous dissipation within the turbulent jet has to be taken into account for such flows.

ACKNOWLEDGMENTS

This research has been supported by a Ph.D. Grant from the French Ministry of Research and Education. The authors gratefully acknowledge Pierre Chardon and Didier Blanc for their valuable contributions to the elaboration of the setup. They would like also to thank Joël Gilbert, Annemie Van Hirtum, Julien Cisonni, and Claire Brutel-Vuilmet for their participation and suggestions on this work.

Agarwal, M. (**2004**). "The false vocal folds and their effect on translaryngeal airflow resistance," Ph.D. thesis, Bowling Green State University, OH.

Agarwal, M., Scherer, R. C., and Hollien, H. (**2003**). "The false vocal folds: Shape and size in frontal view during phonation based on laminagraphic tracings," J. Voice **17**, 97–113.

Bailly, L., Henrich, N., Webb, M., Müller, F., Licht, A. K., and Hess, M. (**2007**). "Exploration of vocal-folds and ventricular-bands interaction in singing using high-speed cinematography and electroglottography," in Proceedings of the 19th International Congress on Acoustics, Madrid, Spain.

Bailly, L., Ruty, N., Van Hirtum, A., Cisonni, J., Pelorson, X., and Henrich, N. (**2006**). "Aerodynamic interaction between the vocal folds and the ventricular bands," in Proceedings of the Seventh Advances in Quantitative Laryngology, Voice and Speech Research, Gröningen, The Netherlands.

Blevins, R. D. (**1992**). *Applied Fluid Dynamics Handbook* (Krieger, Malabar, FL), pp. 71–80.

Erath, B. D., and Plesniak, M. W. (**2006**). "The occurrence of the Coanda effect in pulsatile flow through static models of the human vocal folds," J. Acoust. Soc. Am. **120**, 1000–1011.

Fuks, L., Hammarberg, B., and Sundberg, J. (**1998**). "A self-sustained vocal-ventricular phonation mode: Acoustical, aerodynamic and glottographic evidences," KTH TMH-QPSR **3**, 49–59.

Goldstein, M. (**1976**). *Aeroacoustics* (McGraw-Hill, New York).

Henrich, N. (**2006**). "Mirroring the voice from Garcia to the present day: Some insights into singing voice registers," Logoped. Phoniatr. Vocol. **31**, 3–14.

Henrich, N., d'Alessandro, C., Castellengo, M., and Doval, B. (**2004**). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," J. Acoust. Soc. Am. **115**, 1321–1332.

Henrich, N., Lortat-Jacob, B., Castellengo, M., Bailly, L., and Pelorson, X. (**2006**). "Period-doubling occurrences in singing: The 'bassu' case in traditional Sardinian 'A Tenore' singing," in Proceedings of the Fifth International Conference Voice Physiology and Biomechanics, University of Tokyo, Japan, July.

Hirano, M., Matsuo, K., Kakita, Y., Kawasaki, H., and Kurita, S. (**1983**). "Vibratory behavior versus the structure of the vocal fold," in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, Denver, edited by I. R. Titze and R. C. Scherer, pp. 26–40.

Hirschberg, A., Pelorson, X., Hofmans, G. C. J., Van Hassel, R. R., and Wijnands, A. P. J. (**1996**). "Starting transient of the flow through an in-vitro model of the vocal folds," in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego), pp. 31–46.

Hofmans, G. C. J. (**1998**). "Vortex sounds in confined flows," Ph.D. thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands.

Hofmans, G. C. J., Groot, G., Ranucci, M., Graziani, G., and Hirschberg, A. (**2003**). "Unsteady flow through in-vitro models of the glottis," J. Acoust. Soc. Am. **113**, 1658–1675.

Hollien, H., and Colton, R. H. (**1969**). "Four laminagraphic studies of vocal fold thickness," Folia Phoniatr. **21**, 179–198.

Imagawa, H., Sakakibara, K. I., Tayama, N., and Niimi, S. (**2003**). "The effect of the hypopharyngeal and supra-glottic shapes on the singing voice," in Proceedings of the Stockholm Music Acoustics Conference, Stockholm, Sweden, August.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1267.

Ishizaka, K., and Matsudaira, M. (**1972**). *Fluid Mechanical Considerations of Vocal Cord Vibration, Speech Communications Research Laboratory, Monograph 8* (Santa Barbara, CA), pp. 12–17.

Kucinschi, B. R., Scherer, R. C., DeWitt, K. J., and Ng, T. T. M. (**2006**). "Flow visualization and acoustic consequences of the air moving through a static model of the human larynx," J. Biomech. Eng. **128**, 380–390.

Kundu, P. K. (**1990**). *Fluid Mechanics* (Academic, San Diego).

Lindestad, P. A., Sodersten, M., Merker, B., and Granqvist, S. (**2001**). "Voice source characteristics in Mongolian 'throat singing' studied with high-speed imaging technique, acoustic spectra, and inverse filtering," J. Voice **15**, 78–85.

Lous, N. J. C., Hofmans, G. C. J., Vedlhuis, R. N. J., and Hirschberg, A. (**1998**). "A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design," Acta Acust. **84**, 1135–1150.

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

Miller, J. A., Pereira, J. C., and Thomas, D. W. (**1988**). "Fluid flow through the larynx channel," J. Sound Vib. **121**, 277–290.

Nasri, S., Jasleen, J., Gerratt, B. R., Sercarz, J. A., Wenokur, R., and Berke, G. S. (**1996**). "Ventricular dysphonia: A case of the false vocal fold mucosal travelling wave," Am. J. Otolaryngol. **17**, 427–431.

Pelorson, X., Hirschberg, A., van Hassel, R. R., and Wijnands, A. P. J. (**1994**). "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation: Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.

Pelorson, X., Liljencrants, J., and Kroeger, B. (**1995**). "On the aeroacoustics of voiced sound production," in Proceedings of the 15th International Congress on Acoustics, Trondheim, Norway.

Pinho, S. M. R., Pontes, P. A. L., Gadelha, M. E. C., and Biasi, N. (**1999**). "Vestibular vocal fold behaviour during phonation in unilateral vocal fold paralysis," J. Voice **13**, 36–42.

Roubeau, B., Henrich, N., and Castellengo, M. (**2008**). "Laryngeal vibratory mechanisms: The notion of vocal register revisited," J. Voice, 9 June.

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

Ruty, N., Van Hirtum, A., Pelorson, X., Hirschberg, A., and Lopez-Arteaga, I. (**2006**). "A preliminary study of asymmetric vocal fold vibrations: Modeling and 'in-vitro' validation," in Proceedings of the Seventh International Seminar on Speech Production, Ubatuba, Brazil.

Sakakibara, K. I., Imagawa, H., Niimi, S., and Tayama, N. (**2004**). "Physiological study of the supraglottal structure," in Proceedings of the International Conference on Voice Physiology and Biomechanics, Marseille, France, August.

Sakakibara, K. I., Konishi, T., Kondo, K., Murano, E. Z., Kumada, M., Imagawa, H., and Niimi, S. (**2001**). "Vocal fold and false vocal fold vibrations and synthesis of khoomei," in Proceedings of the International

Computer Music Conference, Havana, Cuba, Sep., pp. 135–138.
Scherer, R. C., Shinwari, D., DeWitt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.
Shadle, C., Barney, A., and Thomas, D. (**1991**). "An investigation into the acoustics and aerodynamics of the larynx," in *Vocal Fold Physiology: Acoustics, Perceptual and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffin and B. Hammarberg (Singular Publishing Co.), pp. 73–82.
Shinwari, D., Scherer, C., DeWitt, K. J., and Afjeh, A. A. (**2003**). "Flow visualization and pressure distributions in a model of the glottis with a symmetric and oblique divergent angle of 10 degrees," J. Acoust. Soc. Am. **113**, 487–497.
Tsai, C., Shau, Y., and Hsiao, T. (**2004**). "False vocal fold surface waves during Sygyt singing: A hypothesis," in Proceedings of the International Conference on Voice Physiology and Biomechanics, Marseille, France, August.
Van Hirtum, A., Cisonni, J., Ruty, N., Pelorson, X., Lopez, I., and Van Uittert, F. (**2007**). "Experimental validation of some issues in lip and vocal folds physical models," Acta Acust. **93**, 314–323.
Wilson, J. E. (**1976**). "Variations of the laryngo-pharynx in singing," The NATS Bulletin **31**, 20–22.
Zhang, Z., Neubauer, J., and Berry, D. (**2006**). "The influence of subglottal acoustics on laboratory models of phonation," J. Acoust. Soc. Am. **120**, 1558–1569.
Zhang, C., Zhao, W., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation, Part II: Effects of flow parameters and ventricular folds," J. Acoust. Soc. Am. **112**, 2147–2154.

Frequency dependence of backscatter from thin, oblique, finite-length cylinders measured with a focused transducer-with applications in cancellous bone

Keith A. Wear[a] and Gerald R. Harris
U. S. Food and Drug Administration, Center for Devices and Radiological Health, Silver Spring, Maryland 20993

(Received 20 May 2008; revised 1 August 2008; accepted 4 August 2008)

A model is presented for the echo from a thin, oblique, finite-length cylinder. The echo is calculated from the line integral of the transducer directivity pattern along the cylinder axis. The model was validated with broadband pulse-echo measurements from (1) a perpendicular (to the ultrasound beam) nylon wire as a function of lateral displacement from the beam center, (2) a tilted nylon wire as a function of the angle of inclination relative to the ultrasound beam, and (3) a quasi-parallel-nylon-wire phantom, which mimicked the scattering properties of cancellous bone. The transducer directivity pattern (as a function of position and frequency) was measured with a membrane hydrophone. The model predicts an approximately cubic frequency dependence of backscatter coefficient from the phantom, as has been observed experimentally in cancellous bone. The model also predicts the relationship between the cylinder length and the exponent of a power law fit to backscatter coefficient versus frequency, which is 4 for very short (compared to a wavelength) cylinders and asymptotically approaches 3 for very long cylinders.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2980524]

PACS number(s): 43.80.Ev, 43.80.Jz, 43.20.Fn [CCC] Pages: 3309–3314

I. INTRODUCTION

Several investigators have proposed models for ultrasonic scattering from human cancellous bone (Wear, 1999; Chaffai *et al.*, 2000; Wear, 2001; Jenson *et al.*, 2003; Padilla *et al.*, 2003). One model, which is based on the theory of Faran (1951) of elastic scattering from cylinders, assumes that the cylinders (e.g., bone trabeculae) are oriented perpendicularly to the ultrasound beam propagation direction (Wear, 1999). This is approximately true in clinical bone sonometry, in which the calcaneus is interrogated in the mediolateral orientation. The model also assumes that the cylinders are longer than the beam width so that potential effects due to finite cylinder lengths may be ignored. Cylinder models have been shown to be useful in many investigations for predicting ultrasonic properties of human calcaneus, including the dependence of backscatter coefficient on frequency (Wear, 1999), the dependence of backscatter coefficient on trabecular thickness (Wear and Laib, 2003), and the dependence of phase velocity on trabecular thickness and scatterer volume fraction (Wear, 2005; Haïat *et al.*, 2007). Figure 1 shows a microcomputed tomography (CT) reconstruction of a slice of human calcaneus. Filamentlike trabeculae, which are apparent throughout the bone, may be thought of as thin cylinders.

In order to broaden the applicability of the cylinder model to bone, a new variation is proposed here that allows for finite length and arbitrary orientation with respect to the ultrasound propagation direction (i.e., not necessarily perpendicular). Unlike the previous model (Wear, 1999), however, the new model requires the cylinders to be thin compared to a wavelength. This is not a serious limitation in clinical bone sonometry, in which the mean trabecular thickness (in human calcaneus) is about 127 μm (Ulrich *et al.*, 1999) while a typical diagnostic wavelength (at 500 kHz) is about 3 mm. Table I compares the present investigation to previous theoretical and experimental investigations of scattering by cylinders. Advantages of the present study relative to many previous studies include applicability to (1) focused transducers (which are preferred for most biomedical applications), (2) arbitrary cylinder length and orientation, (3) backscatter from collections of unresolvable cylinders rather than single, isolated cylinders, and (4) measurements from a cancellous bone mimicking phantom.

II. THEORY

Figure 2 shows the geometry of the measurement. An ultrasound beam propagates along the z axis and scatters off a cylinder centered at a point (x_0', y_0') in the (x_0, y_0) plane.

The pressure field at a point (x_0, y_0) in an observation plane a distance z away from the transducer is given by (Madsen *et al.*, 1984)

$$p(x,y,z;\omega) = T(\omega)B(\omega)A(\omega)D(x_0,y_0,z;\omega), \quad (1)$$

where $T(\omega)$ is the spectrum of the electrical excitation, $B(\omega)$ is the transducer electromechanical conversion factor, $A(\omega)$ is an attenuation compensation function (Oelze and O'Brien, 2002), and $D(x_0,y_0,z;\omega)$ is the directivity, given by the Rayleigh–Sommerfeld diffraction formula [Goodman, 1968, Eqs. (3)–(25)]

[a]Electronic mail: kaw@fda.hhs.gov

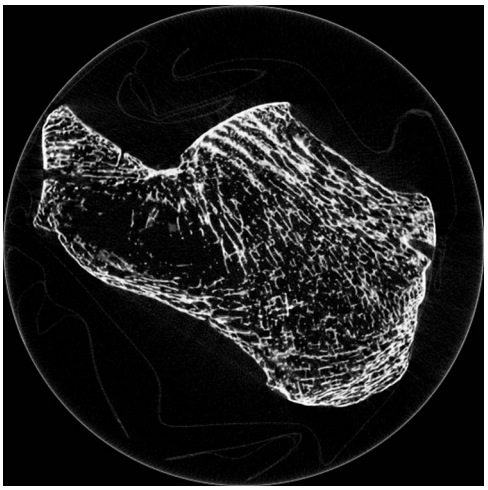


FIG. 1. Micro-CT reconstruction of a slice of human calcaneus. Strandlike trabeculae may be seen to traverse the bone cross section.

$$D(x_0,y_0,z;\omega)=\frac{1}{j\lambda}\int_{\Sigma}\int Q(x_1,y_1)\frac{\exp(jkr_{01})}{r_{01}}d\Sigma, \tag{2}$$

where the transducer surface lies in the (x_1,y_1) plane, $Q(x_1,y_1)$ is an apodization factor, $j=\sqrt{-1}$, λ is the wavelength, $k=2\pi/\lambda$, $\omega=kc$, c is the ultrasound phase velocity, Σ is the transducer surface, $d\Sigma=dx_1dy_1$,

$$r_{01}=\sqrt{(x_0-x_1)^2+(y_0-y_1)^2+z^2},$$

and an obliquity factor has been approximated as 1. For example, in the focal plane of a circular focused transducer with diameter l, $D(x_0,y_0,z;\omega)$ is given by (Goodman, 1968, Sec. 4–2)

$$D(x_0,y_0,z;\omega)=\exp(jkz)\exp\left(j\frac{kr_0^2}{2z}\right)\frac{kl^2}{j8z}\left[2\frac{J_1(klr_0/2z)}{klr_0/2z}\right], \tag{3}$$

where $r_0^2=x_0^2+y_0^2$.

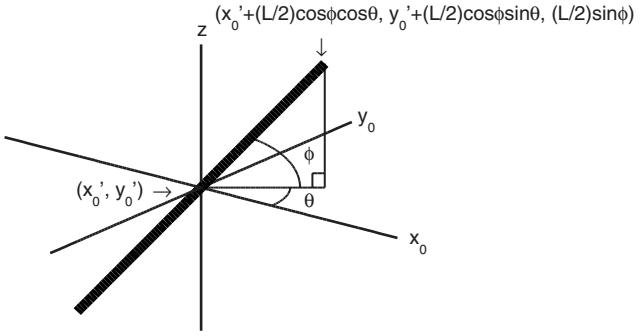


FIG. 2. Measurement geometry. An ultrasound beam propagates along the z axis and scatters off a cylinder centered at a point (x_0',y_0') in the (x_0,y_0) plane. A thin cylinder with length L is centered at (x_0',y_0',z) and inclined at an angle ϕ with respect to the (x_0,y_0) plane and an angle θ with respect to the x_0 axis.

According to the reciprocity theorem, the radiofrequency (rf) echo from a point reflector with reflectivity R at (x_0',y_0',z) interrogated in pulse-echo mode is given by

$$s_P(x_0',y_0',z;\omega)=RT(\omega)B^2(\omega)A^2(\omega)D^2(x_0',y_0',z;\omega). \tag{4}$$

The rf echo from a thin cylinder, with reflectivity R and length L, centered at (x_0',y_0',z) and inclined at an angle ϕ with respect to the (x_0y_0) plane and an angle θ with respect to the x_0 axis (see Fig. 1) is given by the line integral

$$s_C(x_0',y_0',z,\theta,\phi,L;\omega)=RT(\omega)B^2(\omega)A^2(\omega)\int_{-L/2}^{L/2}\times D^2(x_0'+t\cos\phi\cos\theta, y_0'+t\cos\phi\sin\theta, z+t\sin\phi;\omega)dt. \tag{5}$$

Equation (5) ignores a potential angular (ϕ) dependence of the reflectivity R. This assumption is tested in the experiments described below. Generally speaking, Eq. (5) requires knowledge of the directivity pattern throughout the scattering volume. However, a simplified version of Eq. (5) can be

TABLE I. A comparison of theoretical and experimental investigations of scattering by cylinders. In the context of cylinder length, infinite means at least as long as the beam width.

Reference	Transducer focusing properties	Cylinder diameter ($2a$)	Cylinder length (L)	Cylinder orientation	Experimental test target (s)
Faran, 1951	Unfocused	Arbitrary	Infinite	Perpendicular	Single metal cylinders
Lee *et al.*, 1978	Unfocused	Arbitrary	Infinite	Perpendicular	Single metal cylinders, single nylon filament, arrays of nylon cylinders
Ueda and Akita, 1986	Arbitrary	$2a\ll\lambda$	Infinite	Perpendicular	Single stainless steel wire
Stanton, 1988a	Unfocused	Arbitrary	$L\ll\sqrt{z\lambda}$	Perpendicular	Single lead cylinder, single shrimp cylinder
Stanton, 1988b	Unfocused	Arbitrary	$L\ll\sqrt{z\lambda}$	Perpendicular	Single dural and metal cylinders
Li and Ueda, 1990a	Unfocused	Arbitrary	Infinite	Perpendicular	…
Li and Ueda, 1990b	Cylindrical	Arbitrary	Infinite	Perpendicular	Single metal and plastic cylinders
Sheng and Hay, 1993	Unfocused	$2a\ll L$	$L\gg 2a$	Perpendicular	Single stainless steel wires
Stanton, 1993a	Unfocused	$ka\gg 1$	Arbitrary	Perpendicular	Shrimp
Stanton, 1993b	Unfocused	$ka\gg 1$	Arbitrary	Arbitrary	Multiple krill, multiple shrimp
Raum and O'Brien, 1997	Spherical	$2a\ll\lambda$	Infinite	Perpendicular	Single tungsten wires
Wear, 2004	Spherical	Arbitrary	Infinite	Perpendicular	Single nylon wires
Present paper	Arbitrary	$2a\ll\lambda$	Arbitrary	Arbitrary	Single nylon wires, randomly positioned quasi-parallel nylon-wire cancellous bone phantom

applied even when the directivity pattern is characterized in only one plane—e.g., the transducer focal plane—provided the depth dimension (z) of the scattering volume is short compared to the scale of directivity magnitude variation. Then the variation in the magnitude (but not the phase) of the directivity with depth can be ignored throughout the scattering volume so that

$$D(x_0,y_0,z+\Delta z;\omega) \approx D(x_0,y_0,z;\omega)e^{jk\Delta z}. \quad (6)$$

This approximation is tested in the experiments described below. Equation (5) is then

$$s_C(x_0',y_0',z,\theta,\phi,L;\omega) = RT(\omega)B^2(\omega)A^2(\omega)\int_{-L/2}^{L/2} \times D^2(x_0'+t\cos\phi\cos\theta, y_0'+t\cos\phi\sin\theta,z;\omega)e^{j2kt\sin\phi}dt. \quad (7)$$

For a collection of N cylinders distributed throughout the scattering volume, the total received signal power is (assuming incoherent scattering)

$$|S_C(\omega,L)|^2 = R^2T^2(\omega)B^4(\omega)A^4(\omega)\sum_{j=1}^{N}\left|\int_{-L/2}^{L/2} \times D^2(x_{0j}'+t\cos\phi_j\cos\theta_j, y_{0j}'+t\cos\phi_j\sin\theta_j,z;\omega)e^{j2kt\sin\phi_j}dt\right|^2. \quad (8)$$

In this study, the directivity $D(x_0,y_0,z;\omega)$ was measured in water with a hydrophone. A hydrophone measurement at a point (x_0,y_0,z) is given by

$$H(x_0,y_0,z;\omega) = v(\omega)T(\omega)B(\omega)D(x_0,y_0,z;\omega), \quad (9)$$

where $v(\omega)$ is the hydrophone sensitivity.

III. METHODS

A Panametrics (Waltham, MA) transducer (center frequency of 500 kHz, focal length of 1.5 in. and diameter of 1 in.) was used as a source for all measurements and also as a receiver for pulse-echo measurements. This is the same transducer used previously to measure scattering from human cancellous bone samples *in vitro* (Wear, 1999). A Panametrics 5077PR pulser/receiver was used for all measurements.

For beam pattern measurements, the Panametrics transducer was immersed in a water tank. A polyvinylidene fluoride (PVDF) spot-poled membrane hydrophone was used to measure the acoustic pressure field in the transducer focal plane along a radial line perpendicular to the beam axis at 1 mm intervals out to a distance of 1 cm from the beam axis. The PVDF film thickness was 25 μm. The hydrophone incorporated an integral preamplifier (Harris, 1988). The beam pattern was assumed to be radially symmetric so that hydrophone measurements along a single line (that passed through the focal point) completely characterized the transducer output at the focal plane.

The diameter of the hydrophone sensitive element was 0.5 mm, which, due to properties of PVDF film, corresponds to an effective diameter at 500 kHz of 1.8 mm for membrane hydrophones (Robinson, 1991). This diameter is 0.6 times the acoustic wavelength at 500 kHz. Based on Eq. (3) (applied to the hydrophone aperture in this case rather than the transmitting transducer aperture), the angular response of the hydrophone at the maximum measurement angle of 15° is approximately 97% of its response at 0°. (Because of the reciprocity theorem, the receive directivity is the same as the transmit directivity.) Therefore, the effect of the hydrophone angular response on the directivity measurements was negligible.

For pulse-echo measurements on nylon wires and phantoms, the Panametrics transducer was supported by five dimensional (5D) (three linear, two angular) positioning stage. Received rf signals were digitized (8 bits, 10 MHz) using a LeCroy (Chestnut Ridge, NY) 9310C Dual 400 MHz oscilloscope and stored on computer (via general purpose interface bus) for off-line analysis.

The effect of cylinder location within the focal plane was investigated in a water tank as follows. The ultrasound beam propagated in a horizontal direction. A nylon fishing line (diameter=300 μm$\approx\lambda/10$ at 500 kHz) was suspended vertically. A fishing weight was tied at the bottom of the nylon line in order to ensure vertical orientation. The Panametrics transducer was scanned laterally. The variation of the pulse-echo response (viewed on the oscilloscope trace) was found to be symmetric about a central point in a lateral scan, which was designated as $y_0'=0$. The transducer was laterally scanned so that rf data could be acquired at five different lateral positions: $y_0'=0$, 2, 4, 6, and 8 mm. Ten rf data lines were acquired at each position. This process was repeated four times, and the power spectra for each lateral position were averaged.

The effect of the tilt angle between the cylinder and the focal plane was investigated in a water tank as follows. The ultrasound beam propagated in a horizontal direction. A nylon fishing line (again diameter=300 μm$\approx\lambda/10$ at 500 kHz) was stretched horizontally between the two vertical plates that were attached to a horizontal rotational stage. The nylon fishing line was rotated in a horizontal plane. The variation of the pulse-echo response (viewed on the oscilloscope trace) was found to be symmetric about a central angle, which was designated as the 0° tilt angle. rf data were acquired at tilt angles of 0°, 2°, 4°, 6°, and 8°. This process was repeated four times, and the power spectra for each tilt angle were averaged.

Frequency-dependent backscatter coefficients were measured, using methodology described previously (Zagzebski *et al.*, 1993; Wear, 1999), from the following:

(1) a phantom consisting of an array of randomly positioned, quasiparallel, nylon wires (diameter=150 μm$\approx\lambda/20$ at 500 kHz) simulating cancellous bone [mean trabecular diameter in human calcaneus=127 μm (Ulrich *et al.*, 1999)], and
(2) a phantom containing glass beads suspended in agar gelatin (size range: 180–212 μm$\approx\lambda/15$ at 500 kHz, c =1556 m/s, scatterer number density=2/cm^3).



FIG. 3. (Color online) Section from quasiparallel nylon-wire phantom. The nominal wire diameter is 150 μm.

Figure 3 shows a photograph of a portion of the quasiparallel-nylon-wire phantom. In order to characterize frequency dependence of scattering, the experimental backscatter coefficient, $\eta(f)$, for each phantom was fitted to a power law of frequency, $\eta(f)=Af^n$, over the range from 0.3 to 0.7 MHz, which is the same frequency range used previously for cancellous bone (Wear, 1999). Although trabeculae in human calcaneus (Fig. 1) are oriented in a variety of directions, they tend to be inclined at small angles relative to sagittal planes and therefore are approximately perpendicular to the ultrasound propagation direction when the foot is scanned in the mediolateral orientation (perpendicular to Fig. 1) as in clinical bone sonometry. The glass-bead phantom contained Rayleigh scatterers (diameter$\ll\lambda$) that would be expected to yield $n=4$ and therefore provided a useful reference measurement to increase confidence in the backscatter coefficient measurement methodology.

Theoretical values for frequency dependence of backscatter from random distributions of quasi-parallel-nylon wires were obtained by simulations using Eq. (8). The input parameters were $R=1$, $N=100$, $f=0.3$, 0.4, 0.5, 0.6, 0.7 MHz, $L=2,4,6,\ldots$, 20 mm, x_0' and y_0'=uniform random variables ranging between ± 8 mm, and ϕ=uniform random variable ranging between 0° and 15°. Frequency-dependent backscattered power spectral estimates were normalized to the frequency-dependent backscattered power spectrum expected from a random distribution of point scatterers. For each value of L, the simulation was repeated ten times in order to generate means and standard deviations for frequency dependence (n) of backscatter coefficient. The random positions and orientations of the wires were different for each simulation.

IV. RESULTS

Figure 4 shows the effect of cylinder displacement from the focal point. The average power spectra in Fig. 2 are normalized to the $y_0'=0$ power spectrum. The data points correspond to the experimental measurements from the nylon wire. The continuous curves correspond to predictions based on hydrophone measurements and Eq. (7). Fair agreement between the two methods may be seen. The frequency dependence of spectra varies considerably with position throughout the focal plane.

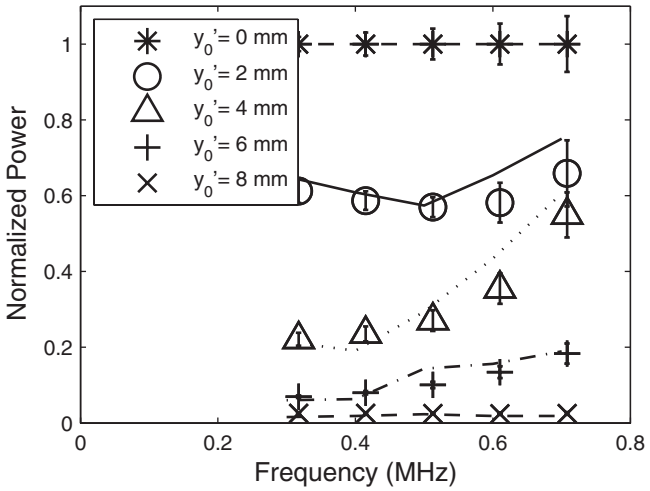


FIG. 4. The effect of cylinder location within the focal plane. Power spectra are normalized to the $y_0'=0$ power spectrum. The individual data points correspond to measurements from the nylon wire. The error bars denote standard deviations. The continuous curves correspond to predictions based on hydrophone measurements and Eq. (7).

Figure 5 shows the effect of tilt angle between the cylinder and the focal plane. The average power spectra in Fig. 3 are normalized to the $\phi=0$ power spectrum. The data points correspond to experimental measurements from the nylon wire. The continuous curves correspond to predictions based on hydrophone measurements and Eq. (7). Fair agreement between the two methods may be seen.

Figure 6 shows the backscatter coefficients measured from the two phantoms. The experimental value for the exponent (n) of frequency dependence for the quasi-parallel-nylon-wire phantom was 2.80 ± 0.25. The value for the exponent (n) for the glass bead phantom was 3.85 ± 0.09, which is close to the asymptotic limit of 4 (the value expected when spherical scatterers are much smaller than the wavelength).

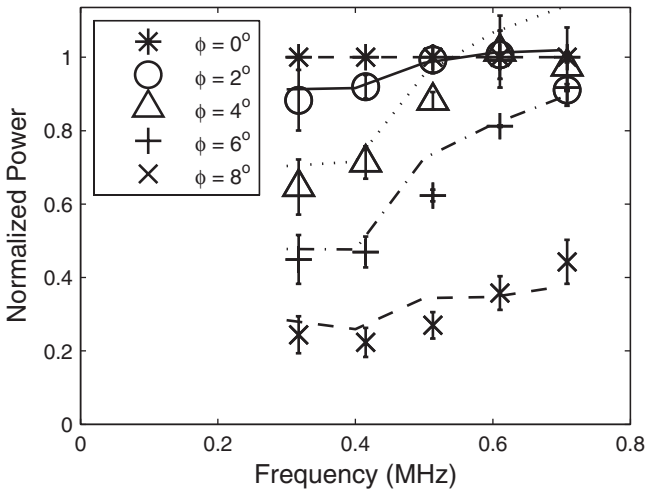


FIG. 5. The effect of angle between the cylinder and the focal plane. Power spectra are normalized to the $\phi=0$ power spectrum. Individual data points correspond to measurements from the nylon wire. The error bars denote standard deviations. The continuous curves correspond to predictions based on hydrophone measurements and Eq. (7).

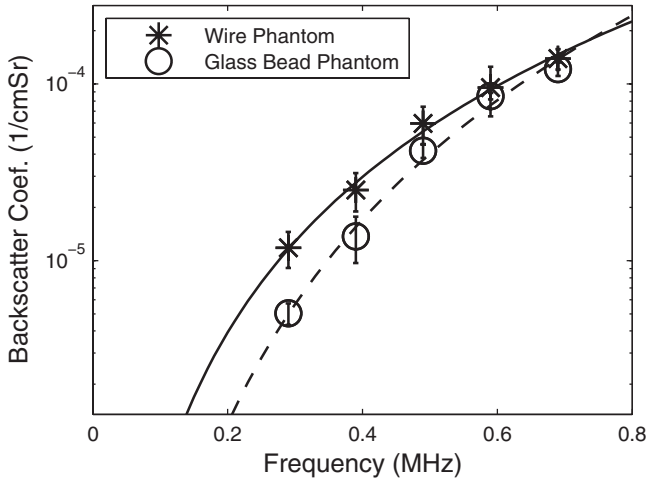


FIG. 6. Backscatter coefficients for the quasi-parallel nylon-wire phantom and for a phantom containing small glass bead scatterers. The lines are power law fits to the data.

Figure 7 shows the estimates of the exponent (n) obtained by simulations based on Eq. (8). In the limit of short L, the exponent approaches 4, as expected when scatterers are much smaller than a wavelength. In the limit of long L, the exponent approaches 3. This is close to the value measured in the quasi-parallel-nylon-wire phantom, 2.80 ± 0.25.

V. DISCUSSION

A. Comparison of the Faran cylinder model with the thin cylinder model

Two limitations of the previously reported Faran cylinder model (Wear, 1999) for scattering from cancellous bone were the assumptions of infinite cylinder length (or at least longer than a beam width) and perpendicular orientation (relative to the ultrasound propagation direction). The present model relaxes these two assumptions, but it is only valid when the cylinder diameter is much smaller than a wavelength (which as mentioned above is true in clinical bone sonometry).

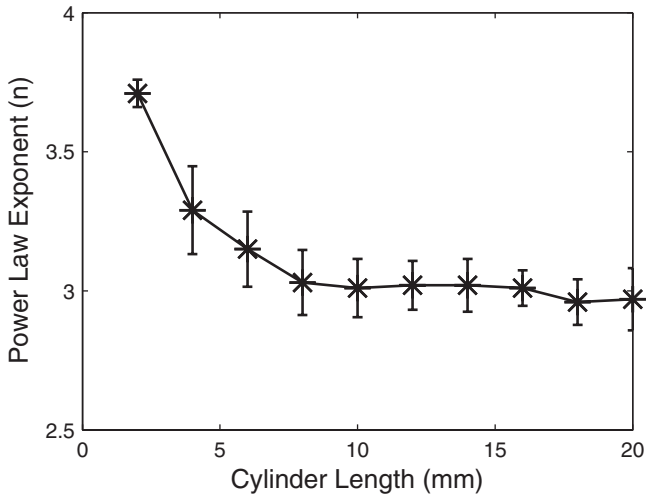


FIG. 7. Estimates of the exponent (n) obtained by simulations based on Eq. (8).

Figure 7 shows that the exponent (n) of frequency dependence varies from 4 for random distributions of very short cylinders (which are effectively point scatterers) to 3 for random distributions of very long cylinders. The exponent is approximately 3 (cubic frequency dependence) for cylinders that are at least 6 mm long. This may help explain why the Faran cylinder model (Wear, 1999) was successful in predicting the frequency dependence of backscatter in human cancellous bone even though it did not take the effects of finite length into account. Apparently (from Fig. 7), the frequency dependence of backscatter (between 0.3 and 0.7 MHz) is essentially the same for all thin cylinders longer than 6 mm. It is plausible that the effective length of dominant trabeculae in human cancellous bone is at least 6 mm. See Fig. 1.

Figure 5 shows that backscatter declines rapidly with angle of orientation. For example, for an 8° inclination, backscattered power at 500 kHz is reduced to 25% of its perpendicular value. This suggests that backscatter from cancellous bone is dominated by those trabeculae that are oriented nearly perpendicular to the ultrasound beam. This may help explain why the Faran cylinder model (Wear, 1999) was successful in predicting the frequency dependence of backscatter in human cancellous bone even though it did not take the effects of nonperpendicularity into account. Apparently (from Fig. 5), trabeculae that are substantially tilted relative to the perpendicular orientation do not produce much backscatter and therefore contribute little to the total backscattered signal.

Figure 5 may also help explain the previously reported result that while the *magnitude* of backscatter is significantly reduced in the anteroposterior (AP) orientation relative to the mediolateral (ML) orientation *in vitro*, the *frequency dependence* of backscatter is similar in both cases (approximately cubic) (Wear, 2000). In both orientations, backscatter is apparently dominated by nearly perpendicular trabeculae that produce approximately cubic-frequency-dependent backscatter, but in the AP orientation, far fewer nearly perpendicular trabeculae are encountered by the ultrasound beam (Wear, 2000).

Because it requires cylinders to be thin relative to the wavelength, the thin cylinder model (unlike the Faran cylinder model) is not useful for estimation of cylinder diameter from pulse-echo measurements. This estimation task would be extremely challenging anyway for cancellous bone at diagnostic frequencies, regardless of model choice, because (1) mean trabecular thickness (in human calcaneus) is about 127 μm (Ulrich *et al.*, 1999) while a typical diagnostic wavelength (at 500 kHz) is about 3 mm, and (2) speckle noise in backscatter signals produces considerable variability in spectral estimates (Wear, 2001). For other estimation tasks, however, when the cylinder diameter is at least on the order of the wavelength, the Faran cylinder model may be preferable.

B. Evaluation of approximations of the thin cylinder model

As stated in Sec. II, Eq. (5) ignores a potential angular (ϕ) dependence of the reflectivity R. In spite of this approxi-

mation, Eq. (5) led to successful prediction of the angular dependence of backscattering from nylon wires. This may be due in part to the rapid decline of backscatter with angle (ϕ) shown in Fig. 5. Apparently, only wires inclined by an angle within about 10° or less to perpendicular (relative to the ultrasound propagation direction) contribute substantial backscatter. It is plausible that the reflectivity R is relatively constant for angles ranging from 0° to 10°, so that the angular variation in the rf echo comes primarily from angular variation in the directivity integral in Eq. (5) (rather than angular variation of R).

As stated in Sec. II, Eq. (6) ignores the variation in the magnitude (but not the phase) of the directivity with propagation distance (z) across the length of a cylinder. This is valid for common ultrasound transducers in the focal zone, where the depth variation of the directivity phase is far more rapid than that the directivity magnitude. Interestingly, Eq. (6) was consistent with experimental measurements even with the relatively tightly focused transducer used here (f 1.5).

VI. CONCLUSION

A model has been presented for the echo from a thin, oblique, finite-length cylinder. The echo is calculated from the line integral of the transducer directivity pattern along the cylinder axis. The model was validated with broadband pulse-echo measurements from (1) a perpendicular nylon wire as a function of lateral displacement from the beam center, (2) a tilted nylon wire as a function of the angle of inclination relative to the ultrasound beam, and (3) a quasi-parallel-nylon-wire phantom, which mimics the scattering properties of cancellous bone. Theory and experiment indicate an approximately cubic frequency dependence of backscatter from the phantom, as has been observed experimentally in cancellous bone.

ACKNOWLEDGMENTS

The author is grateful to Andres Laib, Scanco, Inc., for providing the micro-CT scan in Fig. 1, and to the U.S. Food and Drug Administration Office of Women's Health for providing funding.

Chaffai, S., Roberjot, V., Peyrin, F., Berger, G., and Laugier, P. (**2000**). "Frequency dependence of ultrasonic backscattering in cancellous bone: Autocorrelation model and experimental results," J. Acoust. Soc. Am. **108**, 2403–2411.

Faran, J. J., (**1951**). "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–411.

Goodman, J. W. (**1968**). *Introduction to Fourier Optics* (McGraw-Hill, New York).

Haïat, G., Padilla, F., and Laugier, P. (**2007**). "Analytical modeling of phase velocity dependence on trabecular bone properties: Effect of multiple scattering," in Proceedings of the IEEE Ultrasonics Symposium, New York.

Harris, G. R. (**1988**). "Hydrophone measurements in diagnostic ultrasound fields," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 87–101.

Jenson, F., Padilla, F., and Laugier, P. (**2003**). "Prediction of frequency-dependent ultrasonic backscatter in cancellous bone using statistical weak scattering model," Ultrasound Med. Biol. **29**, 455–464.

Lee, P. P. K., Waag, R. C., and Hunter, L. P. (**1978**). "Swept-frequency diffraction of ultrasound by cylinders and arrays," J. Acoust. Soc. Am. **63**, 600–606.

Li, T., and Ueda, M. (**1990a**). "Sound scattering of a spherical wave incident on a cylinder," J. Acoust. Soc. Am. **87**, 1871–1879.

Li, T., and Ueda, M. (**1990b**). "Analysis of echoes from a cylinder that includes the directivity of the transmitter and receiver," J. Acoust. Soc. Am. **87**, 1880–1879.

Madsen, E. L., Insana, M. F., and Zagzebski, J. A. (**1984**). "Method of data reduction for accurate determination of acoustic backscatter coefficients," J. Acoust. Soc. Am. **76**, 913–1879.

Oelze, M. L., and O'Brien, W. D., Jr. (**2002**). "Frequency-dependent attenuation-compensation functions for ultrasonic signals backscattered from random media," J. Acoust. Soc. Am. **111**, 2308–1879.

Padilla, F., Peyrin, F., and Laugier, P. (**2003**). "Prediction of backscatter coefficient in trabecular bones using a numerical model of three-dimensional microstructure," J. Acoust. Soc. Am. **113**, 1122–1879.

Raum, K., and O'Brien, W. D. (**1997**). "Pulse-echo field distribution measurement technique for high-frequency ultrasound sources," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 810–1879.

Robinson, S. P. (**1991**). in *Output Measurements for Medical Ultrasound*, edited by R. C. Preston (Springer-Verlag, London), p. 64.

Sheng, J., and Hay, A. (**1993**). "Spherical wave backscatter from straight cylinders: Thin-wire standard targets," J. Acoust. Soc. Am. **94**, 2756–1879.

Stanton, T. K. (**1988a**). "Sound scattering by cylinders of finite length. I. Fluid cylinders," J. Acoust. Soc. Am. **83**, 55–63.

Stanton, T. K. (**1988b**). "Sound scattering by cylinders of finite length. II. Elastic cylinders," J. Acoust. Soc. Am. **83**, 64–67.

Stanton, T. K., Clay, C. S., and Chu, D. (**1993a**). "Ray representation of sound scattering by weakly scattering deformed fluid cylinders: Simple physics and application to zooplankton," J. Acoust. Soc. Am. **94**, 3454–3462.

Stanton, T. K.,Chu, D., Wiebe, P. H., and Clay, C. S., (**1993b**). "Average echoes from randomly oriented random-length finite cylinders: Zooplankton models," J. Acoust. Soc. Am. **94**, 3463–3472.

Ueda, M., and Akita, M. (**1986**). "Diffraction correction for backscattering coefficient measurement using echoes scattered by a fine wire," in Proceedings of the 1986 IEEE Ultrasonics Symposium 0090-5607/86/0000-0913.

Ulrich, D., van Rietbergen, B., Laib, A., and Ruegsegger, P. (**1999**). "The ability of three-dimensional structural indices to reflect mechanical aspects of trabecular bone," Bone (N.Y.) **25**, 55–60.

Wear, K. A. (**1999**). "Frequency dependence of ultrasonic backscatter from human trabecular bone: Theory and experiment," J. Acoust. Soc. Am. **106**, 3659–3664.

Wear, K. A. (**2000**). "Anisotropy of ultrasonic backscatter and attenuation from human calcaneus: Implications for relative roles of absorption and scattering in determining attenuation," J. Acoust. Soc. Am. **107**, 3474–3479.

Wear, K. A. (**2001**). "Fundamental precision limitations for measurements of frequency dependence of backscatter: Applications in tissue-mimicking phantoms and trabecular bone," J. Acoust. Soc. Am. **110**, 3275–3282.

Wear, K. A. (**2004**). "Measurement of dependence of backscatter coefficient from cylinders on frequency and diameter using focused transducers—With applications in trabecular bone," J. Acoust. Soc. Am. **115**, 66–72.

Wear, K. A. (**2005**). "The dependences of phase velocity and dispersion on trabecular thickness and spacing in trabecular bone-mimicking phantoms," J. Acoust. Soc. Am. **118**, 1186–1192.

Wear, K. A., and Laib, A. (**2003**). "The relationship between ultrasonic scattering and microarchitecture in human calcaneus," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 979–986.

Zagzebski, J. A., Yao, L. X., Boote, E. J., and Lu, Z. F. (**1993**) in *Ultrasonic Scattering in Biological Tissues*, edited by K. K. Shung and G. A. Thieme (CRC, Boca Raton, FL).

Bowhead whale springtime song off West Greenland

Kathleen M. Stafford
Applied Physics Laboratory, University of Washington (APL-UW), 1013 NE 40th Street, Seattle, Washington 98105

Sue E. Moore
NOAA/Alaska Fisheries Science Center, University of Washington (APL-UW), Seattle, Washington 98105

Kristin L. Laidre
Applied Physics Laboratory, University of Washington (APL-UW), 1013 NE 40th Street, Seattle, Washington 98105 and Greenland Institute of Natural Resources, Box 570, DK-3900 Nuuk, Greenland

M. P. Heide-Jørgensen
Greenland Institute of Natural Resources, Box 570, DK-3900 Nuuk, Greenland

(Received 13 February 2008; revised 30 July 2008; accepted 31 July 2008)

Three songs were recorded from bowhead whales (*Balaena mysticetus*) in Disko Bay, West Greenland, during 59 h of recordings via sonobuoys deployed on seven days between 5 and 14 April 2007. Song elements were defined by units following the protocol of previous description of bowhead whale song. The two most prominent songs were loud, complex, and repeated in long bouts on multiple recording days while the third song was much simpler and recorded on only one day. Bowhead whale simple calls and faint song elements were also recorded using digital audio tape recorders and a dipping hydrophone deployed from the sea ice approximately 100–150 km southwest of Disko Bay on three separate days suggesting that song is also produced in the central portion of Baffin Bay in winter. Songs recorded in Disko Bay are from an area where approximately 85% of the whales have been determined to be adult females. Although it is not known which sex was singing, we speculate that, as in humpback whales (*Megaptera novaeangliae*), male bowhead whales may sing to mediate sexual competition or mate selection behaviors. This is the first detailed description of springtime songs for bowhead whales in the eastern Arctic.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2980443]

PACS number(s): 43.80.Ka, 43.30.Sf [WWA] Pages: 3315–3323

I. INTRODUCTION

Bowhead whales (*Balaena mysticetus*) occur off the coast of West Greenland in early spring, with the population estimated at roughly 1200 (CV=0.3) individuals (Heide-Jørgensen *et al.*, 2007). From February to May, a large proportion of this population concentrates in and near Disko Bay, where they feed intensively on dense layers of epibenthic copepods (Laidre *et al.*, 2007) before migrating across Baffin Bay to the eastern Canadian Arctic in May and June (Heide-Jørgensen *et al.*, 2003). Whales demonstrate highly predictable arrival and departure dates each spring, following observations as far back as the 1700s (Eschricht and Reinhardt, 1861). Satellite tagging studies have revealed strong site fidelity to the region south of Disko Island and have demonstrated that bowhead whales wintering off the west coast of Greenland and summering in the eastern Canadian Arctic likely consists of a single population (Heide-Jørgensen *et al.*, 2003, 2006). Bowheads exhibit considerable sex and age-class segregation across this geographic range, as first noted by commercial whalers (e.g., Southwell, 1898). For example, genetic studies to date show that over 85% of the whales in Disko Bay are adult females, and given they are not accompanied by calves they are either resting or pregnant (Heide-Jørgensen *et al.*, 2007). On the contrary, whales occupying Foxe Basin and northwestern Hudson Bay are nursing females, calves, and subadults (Cosens and Blouw, 2003).

There are few detailed descriptions of calls produced by bowhead whales in the eastern Arctic (Richardson *et al.*, 1995). Conversely, bowhead whale calls have been well described for the western Arctic population (Ljungblad *et al.*, 1980; Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Cummings and Holliday, 1987). Würsig and Clark (1993) compiled results from over 5000 h of recordings obtained from ice-based hydrophone arrays during the 1980s (George *et al.*, 2004) and roughly 500 h of recordings obtained from sonobuoys in the northern Bering, Chukchi and Beaufort Seas (Ljungblad *et al.*, 1986; Würsig *et al.*, 1985). Three types of sounds summarized the acoustic repertoire of bowhead whales in the western Arctic: (1) percussive slaps, blows, gunshot, and crunch sounds; (2) simple frequency-modulated (FM) and complex amplitude-modulated (AM) calls given in no particular order, and (3) long patterned sequences of calls (often called "units" or "notes"), hereafter referred to as song.

In the western Arctic, bowhead songs were usually recorded before mid-May during the initial phase of the spring migration past Barrow, AK (Würsig and Clark, 1993). Songs consisted of one to three themes that include repeated phrases comprised of one to five units. Compared to simple FM and AM calls, which often last only 1–2 s, calls that

comprise songs are longer, exhibit a broader frequency range and much greater FM and AM variations. Songs recorded during the spring migration were usually about a minute long, but song bouts sometimes lasted hours. Usually only one whale of the number within range of the hydrophone array would sing at a time; when that whale stopped, another often began to sing (Cummings and Holliday, 1987). Songs produced by various whales in the same season were similar, but interannual variability in songs was pronounced. For example, distinct bowhead songs were described for each of six springs (1979, 1980, 1982, 1985, 1986, and 1988), with complexity ranging from two to three units lasting a few minutes (Ljungblad *et al.*, 1982), to three to five units, some of which show harmonically unrelated features produced simultaneously and continued for hours (Würsig and Clark, 1993; Fig. 5.9). Singing whales were never observed, so sex, size, and behavior for these animals are unknown.

Despite considerable sampling effort, no records of singing were reported from western Arctic bowhead whales in summer or autumn, causing Würsig and Clark (1993) to speculate that singing is a male breeding display used, as in humpback whales (*Megaptera novaeangliae*), to either attract females or deter rival males (e.g., Cerchio *et al.*, 2001). Modeling estimates based on measurements of fetuses and ovaries from bowheads harvested in Alaska indicate that conception occurs between 2 March and 14 April (Reese *et al.*, 2001). So, songs recorded near Barrow in late April and early May were assumed to be from males singing "late" in the season, resulting in shorter and less frequent songs, again as reported for humpback whales toward the end of the mating season. From autumn recordings of bowhead whales migrating across the Alaskan Beaufort Sea, Blackwell *et al.*, (2007) reported simple or complex calls that were repeated every 2–5 s, for roughly 73 s (range 14 s–5.5 min), with two notable extremes being "call sequences" that lasted 30 and 50 min. Although different from song, in that the call sequences were comprised of recognizable FM and AM calls, this calling behavior exhibited some songlike complexity in calling rate, frequency content, and occasionally included alternating call sequences between at least two whales. Thus, it would appear that bowhead whale calling behavior is more variable than as summarized in Würsig and Clark (1993).

In April 2007, calls and songs of bowhead whales were recorded in Disko Bay, West Greenland. Disko Bay lies at the boundary between the high Arctic waters of Baffin Bay and the sub-Arctic waters of southwest Greenland. Annual sea ice reaches maximum extent in March, and then retreats in April and May such that the bay is ice-free by June. Recordings were made via sonobuoys deployed in open-water areas among drifting sea ice, often near bowheads (2–3 km distances) that were presumably feeding (Laidre *et al.*, 2007). The spectacular call sequences received at the nearby research station are described here.

II. METHODS

A. Recording protocol

Seven ANSSQ-57A type sonobuoys were deployed in 80%–90% ice cover roughly 1 km off the southwest coast of

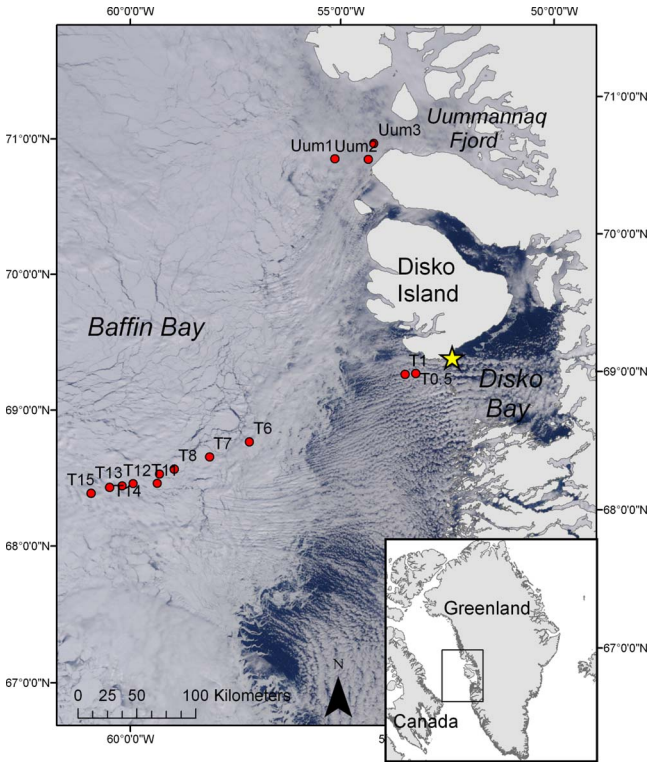


FIG. 1. (Color online) Map of Disko Bay and recording locations in Disko Bay, Uummannaq fjord, and along the hydrographic line. Sonobuoys were deployed <1 km directly offshore (star) from the town of Qeqertarsuaq and signals were received at the Arctic Station, University of Copenhagen. Sea ice conditions from 13 April 2006 were obtained from the MODIS Aqua satellite.

Disko Island, near the town of Qeqertarsuaq, between 5 and 14 April 2007 (Fig. 1, Table I). Sonobuoys are passive acoustic listening systems that contain a hydrophone, signal processing electronics, and a VHF transmitter (Thomas *et al.*, 1986). Signals from the sonobuoy were received by a VHF antenna mounted on the roof of the field station and connected to a laptop computer continuously recording in the laboratory. Hydrophone deployment was set at 122 m depth and all seven instruments recorded sounds from 10 Hz to 20 kHz for roughly 8–9 h. Data were recorded to an external hard drive as wav files for later analysis.

Recordings in sea ice cover >95% were also made at 11 locations that ranged from ~75 to 300 km offshore from Disko Island. These locations were coincident with a line of hydrographic stations and were therefore opportunistic. Recordings were made using a Sony TCD-D7 digital audio tape (DAT) recorder attached to a dipping hydrophone (Offshore Acoustics: frequency response of 10 Hz—40 kHz). Stations were accessed via a Bell 222 helicopter (Fig. 1) and concurrent oceanographic conditions (CTD casts) were made during recordings sessions. Recordings (~12 h) were made after the helicopter engines had shut down by deploying the hydrophone to 10 m depth through an open-water lead or a hole bored in the ice. The total time recorded at each station varied with conditions and battery failure of equipment (temperatures <−25 °C); however, recordings from each station lasted at least 20 min and in some cases up to 45 min.

TABLE I. Summary of acoustic sampling effort via sonobuoys (sb) in Disko Bay and on-ice via hydrophone (h) in Davis Strait and Uummannaq Fjord (Fig. 1).

Date		Start	End	Sample description	Song type recorded
5 April	sb	1720	0200	Bowhead and beluga calls	Faint warble
6 April	sb	1703	0200	Bowhead and beluga calls	Warble, faint screechy
7 April	sb	1643	0120	Bowhead only	Screechy, faint warble, 3+ animals overlapping at times
8 April	sb	1212	2030	Bowhead	Warble and check mark
9 April	sb	1836	0310	Bowhead	Warble, screechy, overlapping multiple animals
10 April	sb	2009	0500	Bowhead	Faint warble, 2+overlapping throughout the recording
14 April	sb	2020	0510	Bowhead	Warble
9 April T7	h	1045	1130	Bowhead	Warble
5 April T8	h	1346	1354	Bowhead and bearded seal	Faint
9 April T11	h	1656	1724	Bowhead and bearded seal	Faint
9 April T12	h	1505	1525	Nothing obvious	
9 April T14	h	1400	1419	Narwhal	Clicks
9 April T15	h	1244	1321	Narwhal	Clicks and cries
14 April U1	h	1335	1442	Narwhal and bearded seal	Clicks and whistles, song
14 April U2	h	1619	1707	Narwhal and bearded seal	Clicks and whistles, song
15 April U3	h	1340	1540	Narwhal and bearded seal	Clicks and whistles, song

B. Data analysis

All 59 h of acoustic data from the sonobuoy recordings were displayed as spectrograms [4096 point fast Fourier transform (FFT), 50% overlap] and visually examined for bowhead whale calls using Raven (Charif *et al.*, 2007). Data were also played back as audio files to help characterize the call types and song notes, units, and themes. Time-frequency measurements were made in Raven (same FFT and overlap as above). The DAT tapes (~6 h) from the dipping hydrophone stations were digitized as 10 min long wav files and were both listened to in real time for bowhead whale sounds and displayed as spectrograms as above.

C. Song classification

Song elements were classified as aurally distinct units (Würsig and Clark, 1993). Each song is composed of multiple units always repeated in the same order. We used onomatopoetic names for the two most common songs as this was more evocative of the differences among them and because a "generic" nomenclature such as "songs I and II" might become confusing if future songs are considerably different. The third song we named the "check mark" song based on spectrographic qualities of the most common unit.

III. RESULTS

The 59 h of recordings via sonobuoys were made on seven days in from 5 to 14 April 2007 (Fig. 1). Of that sample, only about 6 h did not contain any recognizable bowhead whale calls. Fourteen hours on 5–8 April 2007, contained calls that did not overlap others such that parameters of each call could be measured. Songs and calls were often spectacularly loud. Whales were often seen in the area when sonobuoys were deployed, with the closest animals sighted within 2–3 km of the sonobuoy on 6 April 2007. On 9 and 10 April, there were many whales calling and their overlapping songs and calls could not be distinguished from each other; therefore, no measurements were made but song types were noted (Table I). The only other species recorded by sonobuoys were beluga whales (*Delphinapterus leucas*) which were recorded on 5 and 6 April 2007.

A total of approximately 6 h were recorded via dipping hydrophone on three days from 9 to 15 April. Of that sample, bowhead whales were heard on three stations, bearded seals (*Erignathus barbatus*) on two stations, and narwhals (*Monodon monoceros*) on five stations (Table I). The quality of the recordings from Davis Strait was poor mostly due to the noise from people walking on the ice while taking hydrographic measurements, which transmitted well into the water column, but also because below freezing air temperatures (−25 °C) caused the DAT recorder to seize up. Recordings from Uummannaq (Fig. 1) were much better but contained no bowhead whale calls but many vocalizations of bearded seals and narwhals.

A. Song types

Bowhead whales in Disko Bay produced three distinctive songs, each containing two to six units. The first two songs were extremely complex, loud, and showed evidence of two different sounds being produced simultaneously, as was suggested by Würsig and Clark (1993) for western Arctic bowhead whales.

The most common song, named "warble song" for its distinctive concluding unit, consisted of an initial FM upsweep followed without a break by a long (22 s) harmonically rich cry (unit one) that was produced at the same time as two lower frequency units (units 2 and 3), terminating in a long low *warble* (unit 4: Fig. 2). The lowest frequency components (units 2 and 3) were produced simultaneously with unit 1 and occurred in groups of three to seven units with unit 2 followed 0.6 s (±0.25) later by unit 3 which was a

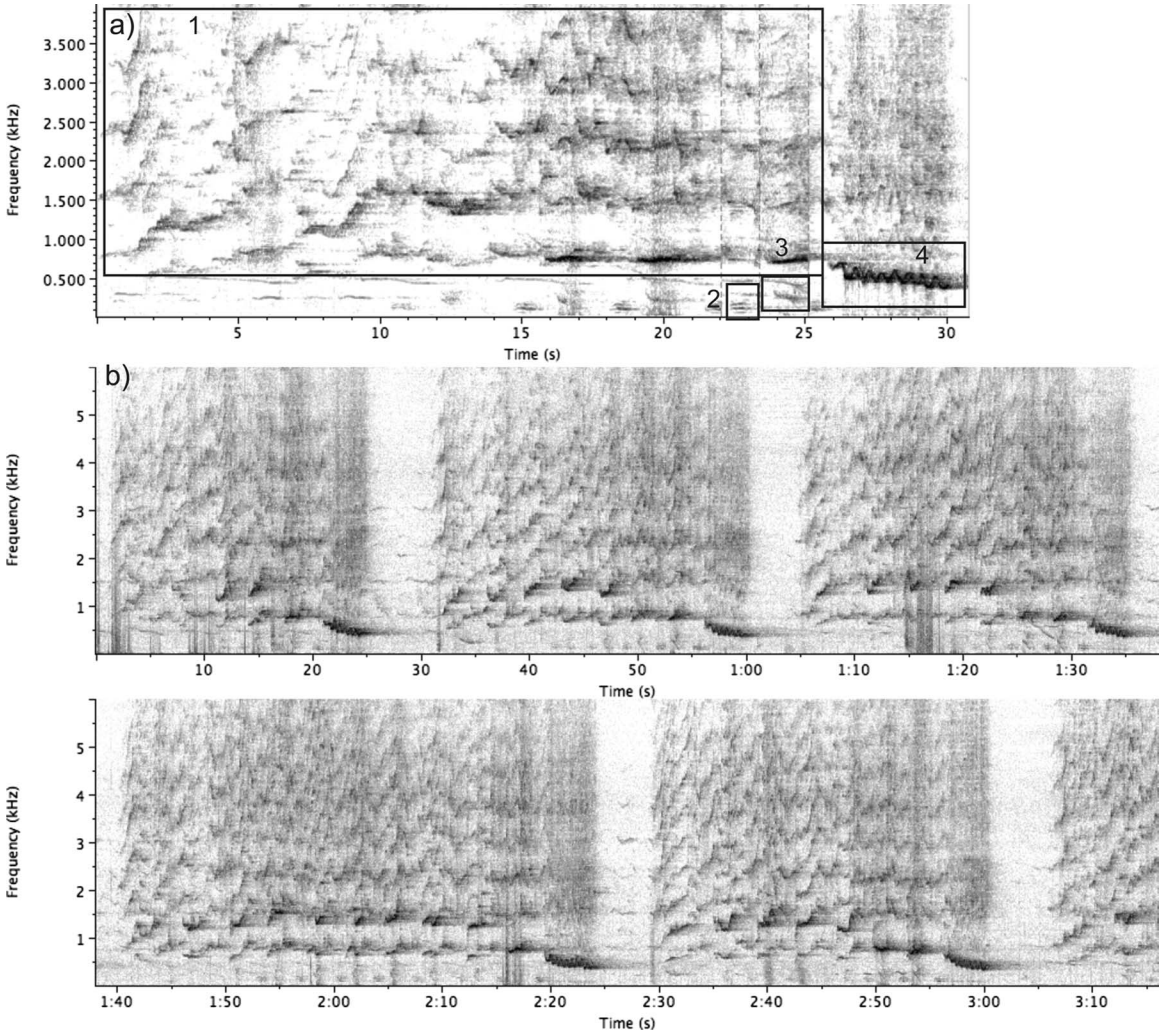


FIG. 2. (a) Warble song units; (b) five warble songs recorded 8 April 2007 (bottom two panels); FFT 4096, 50% overlap, Hanning.

simple 0.7 s (±0.17) down-sweep (Table II). Subsequent unit 2's occurred 1.5 (±0.4) s later. The warble unit lasted almost 5 s with a mean of 7.4 (±1.3) oscillations during each warble. This whole song had a duration of roughly 31 s and was recorded in long repeating bouts during every recording session. The average bout lasted 11.8 ± 9.0 min, with 20 ± 14 songs per bout and the first and last songs in each bout were always shorter than the intervening songs. Although one warble song often dominated the recording due to loudness (likely due to the proximity of the singing whale to the sonobuoy), several other warble songs were often evident in the background, aurally, and on the spectrograms, demonstrating that more than one whale was singing this song at the same time. Our best estimate of the maximum number of animals producing the warble song at any one time was 3, based on overlap of different song parts.

The second most common song, named "screechy song," was also the most complex. The song consisted of six units [Fig. 3(a)], including an aurally distinctive "screeching-squeal" second unit. The song begins with a short low-frequency FM call (unit 1) followed by one to three very *screechy* FM cries (unit 2). This is followed by a loud low-frequency FM down-sweep (unit 3), and then by a series of short, paired, FM cries: first, a high frequency, rich down- or up-down-sweep with tonal elements at ~550 Hz intervals with greatest energy at 1843 Hz that decreased in duration over time (unit 4). The second was a much lower frequency tonal-upsweep (unit 5). Each pair was repeated 3–16 times ($x = 9.2 \pm 2.4$). The song ends with a loud very low-frequency tonal down-sweep (unit 6). The distinctive screechy song starts with a down-sweep (unit 1), which then oscillates one to five times between 1 to 3 kHz before ending in a high pitched *squeal*. The first unit 2 (i.e., screech) was always equal to or longer in duration than subsequent screeches, which never had more than two oscillations.

Screechy songs were recorded on 6, 7, and 9 April, but only songs recorded on 7 April contained distinct units that could be measured. 137 of these songs were audibly distinct on the recordings, 119 contained elements discrete enough to measure, with an additional 8–12 songs visible in spectrograms as background noise. The 137 distinct songs occurred in 19 bouts [Fig. 3(b)]. The mean number of songs/bout was 7.2 (range 1–22). Songs lasted 65.8 ± 11.0 s and bouts lasted 10.3 ± 7.67 min. The intersong interval within bouts was only 8.7 ± 2.4 s. Nearly all (118 of 137) of the songs had two unit 2's (or screeches).

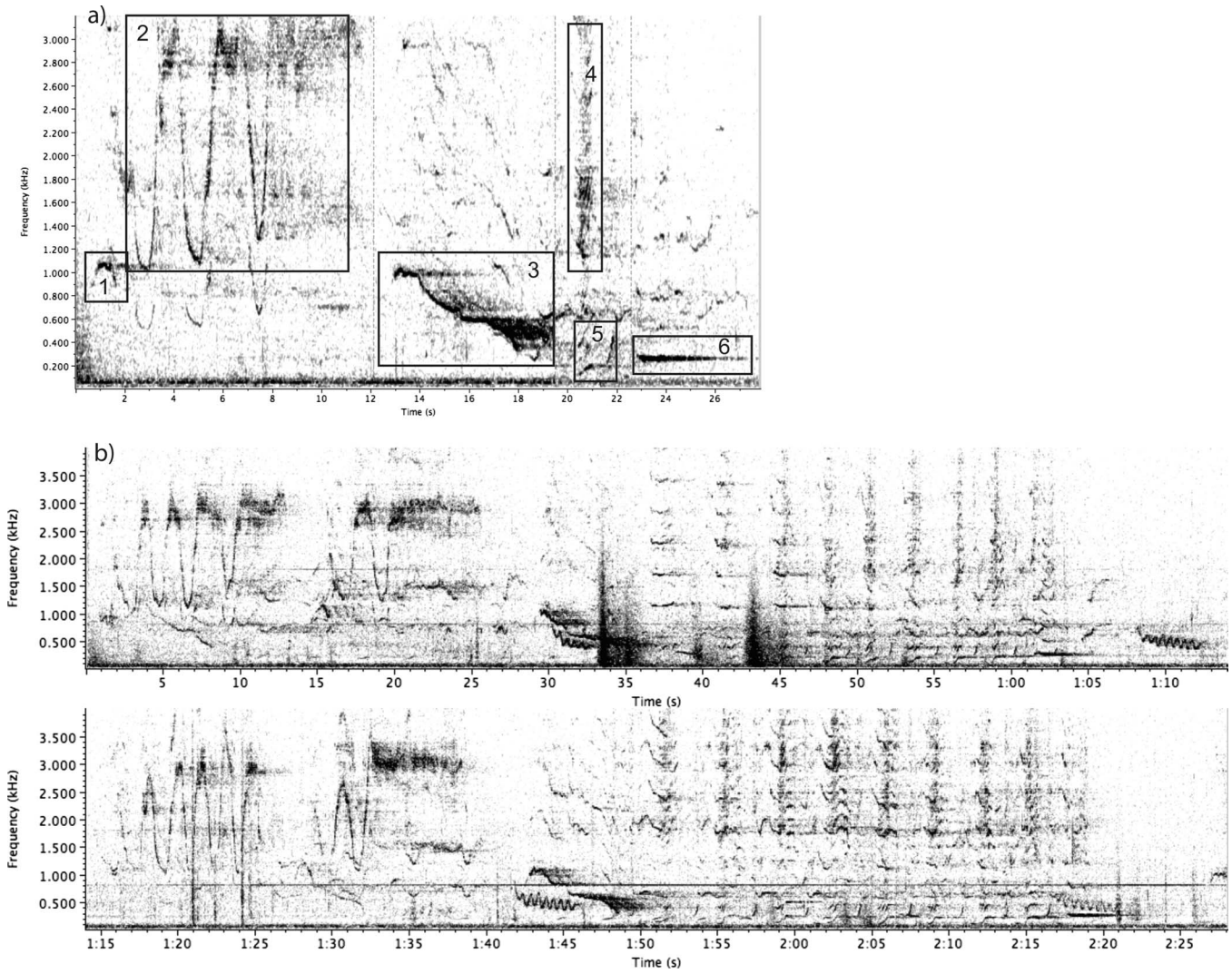


FIG. 3. (a) Screech song units; (b) two screech songs recorded on 7 April 2007 (bottom two panels); FFT 4096, 50% overlap, Hanning window.

The third song (check mark song) was the simplest of the three and was only recorded on 8 April 2007. The check mark song consisted of only two units. The first unit was a down-up swept signal that, on spectrograms, resemble check marks and sounded like "weeoww-whup," which was followed by a single loud tone (Fig. 4). At least 147 of this type

TABLE II. Duration and frequency parameters of song units.

Song	Duration (±s.d.) (s)	No. of song	Start (s) or low (l) freq. (Hz)	End (e) or high (h) freq. (Hz)	Center (c) or Max (m) freq. (Hz)
Warble	30.4 (6.3)				
Unit 1	22.3 (6.2)	1	574.8 (72.0) (l)	5678.9 (298.9) (l)	912.7 (383.9) (m)
Unit 2	4.7 (0.5)	2–7	124.2 (16.8) (s)	81.4 (13.7)	
Unit 3	0.7 (0.2)	1–5	260.9 (39.3) (s)	113.6 (35.7)	
Unit 4	4.76 (0.6)	1–2	664 (44.5) (s)	330 (24.7)	
Screech	65.8 (11.0)				
Unit 1	0.9 (0.2)	1–3	1134.1 (81.0) (s)	823.3 (81.9) (e)	
Unit 2	10.4 (0.9)	1–3	1073.3 (184.5) (l)	3270.1 (197.5) (h)	2280.4 (692.3) (m)
Unit 3	6.4 (0.7)	1	1093.3 (86.0) (s)	349.1 (68.3) (e)	
Unit 4	1.8 (0.5)	3–16	1020.9 (254.3) (l)	4094.9 (585.6) (h)	1843.5 (618.1) (m)
Unit 5	1.8 (0.3)	2–11	147.0 (34.9) (s)	446.3 (34.5) (e)	
Unit 6	3.9 (1.4)	1–4	318.3 (23.2) (s)	203.6 (22.5) (e)	
Check mark	12.2–84.4				
Unit 1	1.6 (0.2)	2–12	1141.5 (103.2) (s)	648.4 (51.8) (e)	
Unit 2	2.1 (0.4)	1			378.7 (13.3) (c)

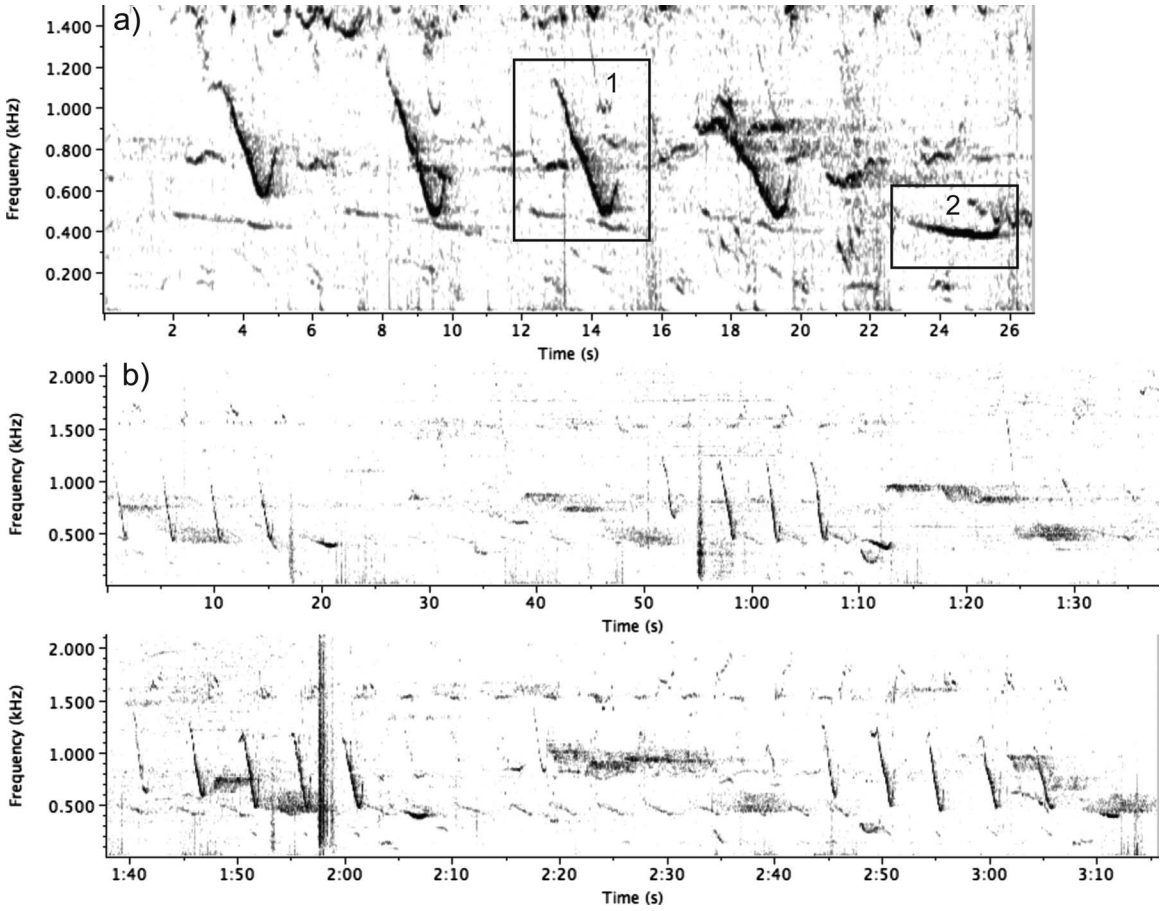


FIG. 4. (a) Check mark song with both units identified; (b) four songs recorded 8 April 2007 (bottom two panels); FFT 4096, 50% overlap, Hanning window.

of song were recorded, although only 107 were distinct enough to measure. The number of check mark units preceding the tone ranged from 2–12 ($x=4.5\pm1.5$). These units lasted 1.6 ± 0.2 s and started, on average, at 1141.5 ± 103.2 Hz ($n=489$). Calls swept down to 433.0 ± 80.7 Hz before the short upsweep to 648.4 ± 51.8 Hz. Not all of these units swept back up, some were simple FM down-sweeps ($n=157$ out of 489). The tone followed all but one of these units and had a center frequency of 378.7 ± 13.3 Hz. Tones lasted 2.1 ± 0.4 s and were the highest amplitude unit. The first unit of each was always highest in frequency and lowest in amplitude. The total duration of a song was a function of the total number of checks and ranged from 12.2 to 84.4 s.

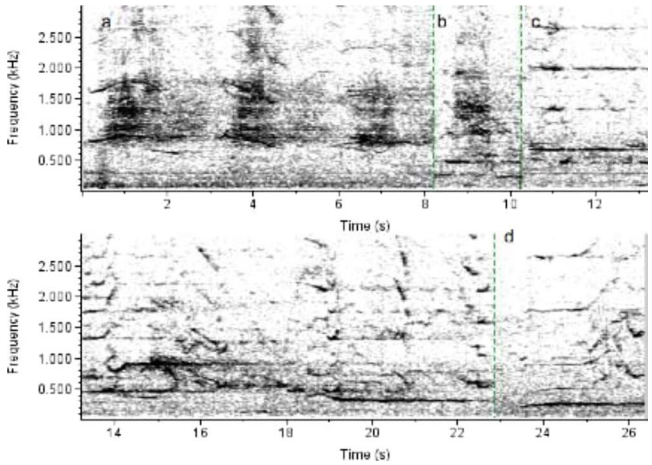


FIG. 5. (Color online) Examples of other nonpatterned bowhead whale calls recorded on sonobuoys in Disko Bay in April 2007. (a) Purr calls; (b) growl call; (c) FM constant and down calls; (d) up call (FFT 2048 points, 50% overlap for all).

B. Geographic range of song

Bowhead whale song elements and simple calls were recorded in the pack ice using dipping hydrophones on 3 of the 11 stations: T7 (warble songs), and simple "up" calls at T8 and T11. The farthest offshore station where bowhead calls where recorded was T11, approximately 300 km (140 nm) from Disko Bay. The most common sound from these pack ice recordings were the distinctive trills of male bearded seals produced as a breeding display (van Parijs and Clark, 2006); narwhal calls were also recorded.

C. Other calls

In addition to the three songs, many other calls were recorded from the bowheads in Disko Bay, with a few examples of spectrograms provided in Fig. 5. These calls did not show any pattern in production and seemed typical of bowhead whale calls recorded elsewhere. Calls recorded included low-frequency (<1 kHz) FM calls labeled "up," "down," "constant," "inflect," and "high" depending on their pattern of modulation and frequency content, and the more

labile AM calls categorized as "growls" (<1 kHz) or "trumpets and screams" (0.5–5 kHz) depending on frequency range (Würsig and Clark, 1993). While this lexicon of seven call types, each lasting only 1–2 s, provides a framework for identifying most nonsong sound production, future sampling and measurement of calls are required to determine if bowheads in Disko Bay produce these calls in a context similar to whales in the western Arctic.

IV. DISCUSSION

This is the first description of bowhead whale song in the eastern Arctic. Previous reports consist only of description of simple calls from a short sampling effort (3 h; 890 calls) in Isabella Bay, Baffin Island, in late summer 1984–1987 (Richardson *et al.*, 1995). Conversely, the calling behavior we describe here consists of patterned sequences of often very loud calls produced by multiple whales. The amplitude and complexity of these songs were so surprising and delightful that we have attempted to describe them in considerable detail that they might be compared with future recordings.

The status of the bowhead whale stock that inhabits the eastern Canadian Arctic, Hudson Bay, and West Greenland is not fully understood, but it seems likely that West Greenland may serve as both a mating and a feeding ground for parts of this population in winter and spring. The animals found in Disko Bay are part of a larger single population that extends into the Canadian high Arctic (Heide-Jørgensen *et al.*, 2007). Disko Bay is one of the most productive regions in West Greenland and offers predictable spring and summer prey resources to a myriad of marine mammals (Heide-Jørgensen and Laidre, 2004). A region such as Disko Bay, where food is seasonally and reliably abundant, allows animals to obtain sufficient nutritional resources and have additional energy to expend on elaborate displays such as those described here.

Singing is a behavioral display usually attributed to males across a broad range of species including insects, amphibians, birds, and whales (Payne and McVay, 1971; Searcy and Andersson, 1986; Romer, 1993). It is unknown if songs are used by cetaceans to advertise a resource (such as food or fitness) or as an intermale spacing mechanism. Humpback whales in a geographic region all sing the same song in the same year with interannual changes in song patterns (Winn *et al.*, 1981). The recording of geographically distinct blue whale (*Balaenoptera musculus*) songs on the same feeding ground has been proposed as a means of segregating different acoustic populations of this species (Stafford, 2003; Mellinger and Clark, 2003) but such population-level acoustic differences are undescribed for bowhead whales despite considerable effort on recordings of bowhead whale acoustics.

The recording of three distinct songs over a brief period in Disko Bay was surprising. Bowhead whales from the Bering Sea stock have been described as changing song between years, but all singers in one year sang the same song (Würsig and Clark, 1993). Overlapping songs recorded simultaneously in Disko Bay show that several whales sang the same song at the same time so the songs were not unique to individuals and different songs were sung at the same time so songs were not unique to certain time periods. Similar to other reports of song, however, songs described here lasted roughly 1 min with song bouts continuing for hours and all ended with loud notes. Although we have no evidence that it is only male bowhead whales that sing, we suggest that this may be the case based on evidence from other species of large whales. Based on DAT recordings in the pack ice, bowhead whales appeared to be singing over a larger area than just Disko Bay; however, song elements and calls were not as loud as those recorded in Disko Bay and no bowhead whales were seen at, nor enroute to, the recording site, suggesting that singing whales were at least 10 s of kilometers from the recording sites. Assuming similar source and ambient noise levels as bowhead whales in the Beaufort Sea, it is unlikely that whales in Disko Bay were those recorded in the pack ice. Bowhead whale song has only been estimated to be detected out to ~11 km (Cummings and Holliday, 1987; Clark *et al.*, 1996).

Several factors may explain why bowhead whale singing is so pronounced in Disko Bay compared to Alaska. First, the recordings were completed somewhat earlier in the season (early April) and prior to migration compared to the typical sampling season during migration in the Beaufort Sea (mid-April to early May; Clark and Johnson, 1984; Cummings and Holliday, 1987). Second, early spring access to bowhead whales offshore West Greenland is logistically less challenging than in the Bering Sea and this aided in obtaining recordings from whales prior to migration. Finally, singing might also be more intense if Disko Bay is an important mating ground. Studies in the Bering Sea suggest that conception occurs between 2 March and 14 April, coinciding well with the period when the whales are in Disko Bay and the period the songs were recorded here (Reese *et al.*, 2001). Genetic studies from biopsy dart samples demonstrate that approximately 85% of the whales in Disko Bay in spring 2007 (n =157) were adult females and observations during spring 2000–2007 have shown that it is primarily mature female whales (lengths over 14 m) that occur offshore West Greenland. Genetic results suggest that the fraction of adult males is low and small (presumably young) whales are rarely seen. There are only a few records of newborns from the whaling period in the 19th century [Eschricht and Reinhardt 1861; Greenland Institute of Natural Resources (unpublished data)].

Seasonally varying hormone levels in males clearly influence singing behavior in male birds (Arnold, 1982) and by extension have been inferred in singing behavior of male whales (Cato *et al.*, 2001). If singing behavior is hormonally driven in bowhead whales singing in early April but seldom in May could reflect decreasing hormone levels. It could also represent the presence of fewer animals as bowhead whales appear to migrate out of Disko Bay in mid-May (Heide-Jørgensen *et al.*, 2006). The fact that almost no records of females with newborns have been detected in Disko Bay supports the hypothesis that this area is a mating ground where intensive singing activity is a part of the mating strategy. The fraction of males, based on biopsy results, that have been detected in Disko Bay is low [<15%, Greenland Institute of Natural Resources (unpublished data)]. This may be a

region where large reproductive males have somehow actively excluded other males, a behavior seen in taxa from frogs to elephants (Wells, 1977; Rasmussen *et al.*, 2008) or simply sexual segregation on a breeding ground, similar to that found in sperm whales (Whitehead *et al.*, 1989). The abundance estimate for bowhead whales in West Greenland has increased recently (Heide-Jørgensen *et al.*, 2007), and this in combination with some cyclical reproductive activity could explain why Disko Bay may have only recently become a mating ground with intensive singing activity.

The three songs we documented over a week were all distinctly different from one another, with both the warble and screech songs dominating the data throughout the entire recording period. We are less convinced that the check mark song is a form of true "song" because it is so simple relative to the other two songs and shows similarities to the repeated series of simple calls described as "call sequences" by Blackwell *et al.* (2007). Nevertheless, it stood out among the general cacophony of the Disko Bay recordings. Clearly the next step in acoustic studies of Disko Bay is to obtain more recordings during the same time frame over multiple years to see if these songs change over time. Longer-term recordings would define the seasonal and diel occurrence of singing behavior in Disko Bay. Finally, intra- and interannual broad scale comparisons of acoustic behavior of bowhead whales from multiple sites (for example, the northern Bering Sea, the Beaufort Sea, and Spitsbergen) would permit broad-scale geographic variations in acoustic repertoires to be studied.

ACKNOWLEDGMENTS

This study was funded by the NOAA Ocean Exploration grant to K.L.L. and the Greenland Institute of Natural Resources. The Arctic Station of the University of Copenhagen provided the field base and Air Greenland provided helicopter support. We thank the hunters of Qeqertarsuaq and Qaanaaq for help with the field operations.

Arnold, A. P. (**1982**). in *Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Academic, New York), pp. 75–94.

Blackwell, S. B., Richardson, W. J., Greene, C. R., Jr., and Streever, B. (**2007**). "Bowhead whale (*Balaena mysticetus*) migration and calling behavior in the Alaskan Beaufort Sea, autumn 2001-04: an acoustic localization study," Arctic **60**, 255–270.

Cato, D. H., Paterson, R. A., and Paterson, P. (**2001**). "Vocalization rates of migrating humpback whales over 14 years," Mem. Queensland Mus. **47**, 481–489.

Cerchio, S., Jacobsen, J. K., and Norris, T. F. (**2001**). "Temporal and geographical variation in songs of humpback whales, *Megaptera novaeangliae*: synchronous change in Hawaiian and Mexican breeding assemblages," Anim. Behav. **62**, 313–329.

Charif, R. A., Clark, C. W., and Fristrup, K. M. (**2007**). *Raven Pro 1.3 User's Manual* (Cornell Laboratory of Ornithology, Ithaca, NY).

Clark, C. W., Charif, R. A., Mitchell, S., and Colby, J. (**1996**). "Distribution and behavior of the bowhead whale, *balaena mysticetus*, based on analysis of acoustic data collected during the 1993 spring migration off Point Barrow, Alaska," Rep. Int. Whal. Comm. **46**, 541–552.

Clark, C. W., and Johnson, J. H. (**1984**). "The sounds of the bowhead whales, *Balaena mysticetus*, during the spring migrations of 1979 and 1980," Can. J. Zool. **62**, 1436–1441.

Cosens, S. E., and Blouw, A. (**2003**). "Size- and age-class segregation of bowhead whales summering in northern Foxe Basin: A photogrammetric analysis," Marine Mammal Sci. **19**, 284–296.

Cummings, W. C., and Holliday, D. V. (**1987**). "Sounds and source levels from bowhead whales off Pt. Barrow, Alaska," J. Acoust. Soc. Am. **82**, 814–821.

Eschricht, D. F., and Reinhardt, J. (**1861**). "On the Greenland Right-Whale (*Balaena mysticetus L.*) with special reference to its geographic distribution and migrations in times past and present (in Danish). Om Nordhvalen (*Balaena mysticetus L.*) navnlig med Hensyn til dens Udbredning i Fortiden og Nutiden og til dens ydre og indre Saerkjender," K. Danske Videnskabernes Selskabs Skrifter. Series 5, naturvidenskabelig og mathematisk Afdeling **5**, 433–590.

George, J. C., Zeh, J., Suydam, R., and Clark, C. W. (**2004**). "Abundance and population trend (1978–2001) of western Arctic bowhead whales surveyed near Barrow, Alaska," Marine Mammal Sci. **20**, 755–773.

Heide-Jørgensen, M. P., and Laidre, K. L. (**2004**). "Declining open water refugia for top predators in Baffin Bay and adjacent waters," Ambio **33**(8), 488–495.

Heide-Jørgensen, M. P., Laidre, K. L., Borchers, D., and Samarra, F. I. P. (**2007**). "Increasing abundance of bowhead whales in West Greenland," Biol. Lett. **3**, 577–580.

Heide-Jørgensen, M. P., Laidre, K. L., Jensen, M. V., Dueck, L., and Postma, L. D. (**2006**). "Dissolving stock discreteness with satellite tracking: Bowhead whales in Baffin Bay," Marine Mammal Sci. **22**, 34–45.

Heide-Jørgensen, M. P., Laidre, K. L., Wiig, O., Jensen, M. V., Dueck, L., Maiers, L., Schmidt, H. C., and Hobbs, R. C. (**2003**). "From Greenland to Canada in two weeks: Movements of bowhead whales, *Balaena mysticetus*, in Baffin Bay," Arctic **56**, 21–31.

Laidre, K. L., Heide-Jørgensen, M. P., and Nielsen, T. G. (**2007**). "The role of the bowhead whale as a predator in West Greenland," Mar. Ecol.: Prog. Ser. **346**, 285–297.

Ljungblad, D. K., Leatherwood, S., and Dahlheim, M. E. (**1980**). "Sounds recorded in the presence of an adult and calf bowhead whales," Mar. Fish. Rev. **42**, 86–87.

Ljungblad, D. K., Moore, S. E., and VanSchoik, D. R. (**1986**). "Seasonal patterns of distribution, abundance, migration and behavior of the western Arctic stock of bowhead whales, *Balaena mysticetus*, in Alaskan Seas," Rep. Int. Whal. Comm. **8**, 177–205.

Ljungblad, D. K., Thompson, P. O., and Moore, S. E. (**1982**). "Underwater sounds recorded from migrating bowhead whales, *Balaena mysticetus*, in 1979," J. Acoust. Soc. Am. **71**, 477–482.

Mellinger, D. K., and Clark, C. W. (**2003**). "Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic," J. Acoust. Soc. Am. **114**, 1108–1119.

Payne, R. S., and McVay, S. (**1971**). "Songs of humpback whales," Science **173**, 585–597.

Rasmussen, H. B., Okello, J. B. A., Wittemyer, G., Siegismund, H. R., Arctander, P., Vollrath, F., and Douglas-Hamilton, I. (**2008**). "Age- and tactic-related paternity success in male African elephants," Behav. Ecol. **19**, 9–15.

Reese, C., Calvin, J. A., George, J. C., and Tarpley, R. J. (**2001**). "Estimation of fetal growth and gestation in bowhead whales," J. Am. Stat. Assoc. **96**, 915–938.

Richardson, W. J., Finley, K. J., Miller, G. W., Davis, R. A., and Koski, W. R. (**1995**). "Feeding, social and migration behavior of bowhead whales, *Balaena mysticetus*, in Baffin Bay vs the Beaufort Sea – regions with different amounts of human activity," Marine Mammal Sci. **11**, 1–45.

Romer, H. (**1993**). "Environmental and Biological Constraints for the Evolution of Long-Range Signaling and Hearing in Acoustic Insects," Philos. Trans. R. Soc. London, Ser. B **340**, 179–185.

Searcy, W. A., and Andersson, M. (**1986**). "Sexual selection and the evolution of song," Annu. Rev. Ecol. Syst. **17**, 507–533.

Southwell, T. (**1892**). "The migration of the Right Whale (*Balaena mysticetus*)," Natural Science **12**, 397–414.

Stafford, K. M. (**2003**). "Two types of blue whale calls recorded in the Gulf of Alaska," Marine Mammal Sci. **19**, 682–693.

Thomas, J. A., Fisher, S. R., and Awbrey, F. A. (**1986**). "Acoustic techniques in studying whale behavior," Rep. Int. Whal. Comm. **8**, 121–138.

van Parijs, S. M., and Clark, C. W. (**2006**). "Long-term mating tactics in an aquatic-mating pinniped, the bearded seal, *Erignathus barbatus*," Anim. Behav. **72**, 1269–1277.

Wells, K. D. (**1977**). "Territoriality and male mating success in the green frog (*Rana clamitans*)," Ecology **58**, 750–762.

Whitehead, H., Weilgart, L., and Waters, S. (**1989**). "Seasonality of sperm whales off the Galapagos Islands, Ecuador," Rep. Int. Whal. Comm. **39**, 207–210.

Winn, H. E., Thompson, T. J., Cummings, W. C., Hain, J., Hudnall, J., Hays,

H., and Steiner, W. W. (**1981**). "Song of the humpback whale – population comparisons," Behav. Ecol. **8**, 41–46.
Würsig, B., and Clark, C. (**1993**). *The Bowhead Whale*. edited by J. J. Burns, J. J. Montague, and C. J. Cowles (Society for Marine Mammalogy, Allen, Lawrence, KS), Special Publicatoin No. 2, pp. 157–199.
Würsig, B., Dorsey, E. M., Fraker, M. A., Payne, R. S., and Richardson, W. J. (**1985**). "Behavior of bowhead whales, *balaena mysticetus*, summering in the Beaufort Sea: A description," Fish. Bull. **83**, 357–377.

Beamwidth control and angular target detection in an echolocating bottlenose dolphin (*Tursiops truncatus*)

Patrick W. Moore
SPAWARSYSCEN-San Diego, Code 2351, 53560 Hull Street, San Diego, California 92152-5001

Lois A. Dankiewicz
Science Applications International Corporation, 4065 Hancock Street, Mailstop Q1-A, San Diego, California 92110

Dorian S. Houser
BIOMIMETICA, 7951 Shantung Drive, Santee, California 92071

(Received 17 January 2008; revised 23 July 2008; accepted 31 July 2008)

Bottlenose dolphin (*Tursiops truncatus*) echolocation beams are typically characterized as symmetrical −3 dB beamwidths; however, the functional width of the beam during target detection has not been explored. Angular target detection thresholds of an echolocating dolphin were examined to more fully describe the functional characteristics of the echolocation beam. The dolphin performed an echolocation detection task with its head held in a fixed orientation. Targets were placed 9 m in front of the dolphin [0° *position* (P_0)] and systematically moved right or left until target detection reached chance probability. A 24-element hydrophone array placed 1 m in front of the dolphin was used to measure vertical and horizontal echolocation beamwidths. Detection thresholds were 26° left and 21° right of P_0 for a sphere target and 19° left and 13° right of P_0 for a cylinder target. Estimates of maximum horizontal and vertical beamwidths ranged up to 40° and 29°, respectively, and exhibited large variability. The dolphin nominally steered the maximum response axis of the echolocation beam up to 18° in the horizontal, 12° in the upward vertical, and 4° in the downward vertical. These results suggest that the dolphin can steer and modify the width of the echolocation beam. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2980453]

PACS number(s): 43.80.Ka, 43.80.Jz [WWA] Pages: 3324–3332

I. INTRODUCTION

Dolphin biosonar (echolocation) is characterized by the transmission of pulsed signals, or clicks, for use in underwater navigation and foraging. Echolocation clicks, which are widely believed to be generated by a pair of phonic lips associated with the nasal passages, are short in duration (~100 μs) and broadband (−3 dB bandwidths potentially exceeding 85 kHz; Aroyan *et al.*, 1992; Cranford, 1992; Houser *et al.*, 1999). The echolocation beam of the bottlenose dolphin has been characterized as highly directional (Evans, 1973; Au *et al.*, 1978; Au, 1980; Pilleri, 1982; Au *et al.*, 1986). Past measurements of the dolphin biosonar transmission beam show, on average, −3 dB beamwidths of approximately 10° in both the vertical and horizontal planes and a maximum response axis (MRA) directed forward from the melon and elevated ~5° above the plane of the supposed click source (Au *et al.*, 1986). Estimates of beluga (*Delphinapterus leucas*) emitted beams showed similar findings (Au *et al.*, 1987).

Estimates of the echolocation beam pattern have been attempted on several species of free-ranging odontocetes. For example, using a four-element planar hydrophone array, echolocation clicks from free-ranging white-beaked dolphins have been recorded and used to estimate −3 and −10 dB beams of 8° and 10°, respectively (assuming that the sound beam was cylindrical and symmetrical; Rasmussen *et al.*, 2004). Similar approaches have been applied to the harbor porpoise and sperm whale (Au *et al.*, 1999; Møhl *et al.*, 2003) to estimate or approximate echolocation beam patterns. Other approaches, not using planar arrays, have also been attempted; e.g., a dorsally mounted two-element data logger was used to estimate the echolocation beam pattern of the finless porpoise (Akamatsu *et al.*, 2005) by exploiting echolocation signals reflected from the water surface. Measuring the emitted beams of free-swimming animals can reveal the physical properties of the sound field and how those properties contribute to proximate foraging strategies and communal foraging. However, this information alone is insufficient unless it is coupled with an understanding of the limits on the spatial use of the biosonar field, i.e., what geometric space in front of the dolphin contains echoic information that can be used for purposes of foraging and navigation.

Through controlled experiments, focused study of object detection and discrimination has been possible in the bottlenose dolphin. It has been shown that dolphins can use biosonar to detect the presence of small stainless steel spheres (7.62 cm diameter) at distances up to 113 m (Au and Snyder, 1980). Dolphins can also discriminate between objects differing in size, structure, shape, and material composition (Nachtigall, 1980). They have demonstrated the ability to discriminate between aluminum, copper, and brass circular targets; between circles, squares, and triangular targets; and can recognize aspect-dependent objects at random orientations (Au and Snyder, 1980; Hammer and Au, 1980; Au and Pawloski, 1992). Most of these studies were conducted with

 © 2008 Acoustical Society of America

targets that were intentionally placed a set distance away, directly in front of the animal, presumably on or near the MRA of the echolocation beam. However, a recent study of echolocation in free-swimming bottlenose dolphins suggested an ability to detect and discriminate targets off of the presumed MRA (Houser *et al.*, 2005; Martin *et al.*, 2005). In that study, dolphins carried a biosonar measurement tool (BMT) in their mouths while actively searching for proud (bottom resting) targets in a bay environment. The BMT collected echolocation clicks 1 m in front of the animal and, therefore, on the presumed MRA of the echolocation beam (Houser *et al.*, 2005). Simultaneously, the dive depth and orientation of the dolphin were also recorded. Comparisons of animal orientation relative to the known location of targets suggested that the dolphin often initially detected the target off of the MRA (off-MRA) and then oriented toward the target for final echo inspection and discrimination.

Capitalizing on these prior observations of peripheral (presumed off-MRA) target detection, we examined the ability of the dolphin to detect sphere and cylinder targets as they were placed at progressively greater distances to the left or right of a point (P_0) directly in front of the dolphin. Detection thresholds, expressed as angular displacement from P_0, were determined for each of the targets to the left and right of P_0. Using a 24-element hydrophone array, several properties of the echolocation signals were measured, including click-by-click measures of the beamwidth in the horizontal and vertical planes. The results were used to define target-specific limits on the use of the echolocation beam and to describe voluntary control of the echolocation beam in the bottlenose dolphin.

II. METHODS

A. Subject

A 37 yr old male bottlenose dolphin participated in the study. The dolphin was housed in a floating pen complex in San Diego Bay at the Space and Naval Warfare Systems Center (SSC SD). Hearing tests conducted on the subject four years prior to this study showed that the subject had hearing loss above 55 kHz and was less sensitive (~16–33 dB) in his right ear than his left across the frequency range of 10–40 kHz (Brill *et al.*, 2001). This experiment was approved by the Biosciences Division, Institutional Animal Care and Utilization Committee, and followed all applicable U.S. Department of Defense guidelines for the use and treatment of animals in science.

B. Stimuli and apparatus

The dolphin was trained to station on a biteplate and echolocate to detect targets. Testing was conducted from a single netted pen enclosure (9.1×18.3 m^2; ~10.5 m deep) within San Diego Bay. Targets used in the detection task consisted of a 17.4 cm diameter water-filled stainless steel sphere (SPH) and a hollow aluminum cylinder (17.8 cm height, 7.6 cm diameter, 9.5 mm wall thickness; CYL). For SPH and CYL, the measured target strengths (TSs) were −27 and −18 dB, respectively.



FIG. 1. (Color online) Biteplate apparatus used to ensure immobility of the dolphin during the experiment. Located directly below the biteplate is a set of receive hydrophones used to collect target echoes (data to be presented in a subsequent paper).

The subject was required to station on a submerged biteplate for the echolocation task (Fig. 1). During training and testing phases of the experiment, the targets were positioned 2.5 m below the water surface and 9.14 m in front of the biteplate (Fig. 2). The biteplate was suspended 1.3 m below the water surface and was attached to a rigid aluminum pole. During training sessions, various angles of downward tilt could be obtained by pivoting the biteplate at the attachment point to the pole. A response paddle was located 0.6 m to the right and slightly behind the biteplate. To ensure that pretrial echolocation on the target was prevented, an acoustically opaque screen (1.1×0.9 m^2, 12.7 mm sheet polyvinyl chloride (PVC) coated with 6.3 mm closed cell neoprene) was situated 23 cm beyond the biteplate. The screen was attached to a pulley system so that it could be lifted out of the water at the beginning of each trial. A larger sheet of visually opaque polyethylene (0.4 mm thick) attached to a water-flooded 1.27 cm PVC pipe frame was fixed directly behind the acoustically opaque screen. This sheet prevented visual inspection of the field but was acoustically transparent and

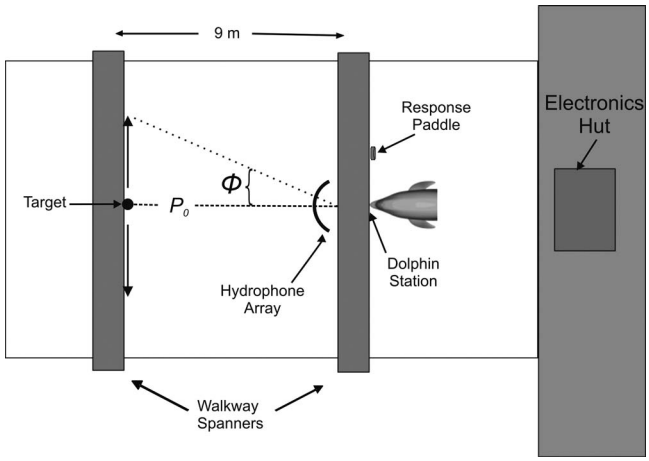


FIG. 2. Experimental setup: the shaded rectangles correspond to spanners placed across the pen and from which the equipment was staged. Targets placed at P_0 were 9 m away from the dolphin. The hydrophone array was placed 1 m in front of the location of the dolphin's phonic lips. The response paddle was placed to the right of the biteplate stationing apparatus. Targets were progressively moved either to the left or to the right. The angle of displacement (Φ) at the point of a 50% detection probability was used to define the angular detection threshold.

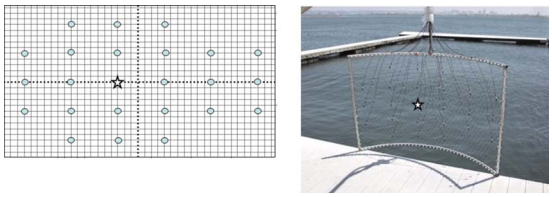


FIG. 3. (Color online) The 24-element hydrophone array used to measure the beam pattern of the dolphin during target detection trials. The left graphic shows a planar display of the array arc (shown on the right). The star within the middle of the graphic denotes the P_0 hydrophone, the hydrophone that was aligned with the main axis of the dolphin to the target when the target was placed directly in front of the dolphin at P_0.

the frame extended beyond the maximum angular displacement of the target, thus allowing echolocation signals to pass without interference from the PVC frame.

An array of 24 omnidirectional hydrophones (Reson TC4013) was attached to a monofilament web mesh placed 0.9 m from the dolphin's rostrum, equating to 1.2 m from the presumed source of the echolocation clicks (Cranford *et al.*, 1996; Cranford, 2000). The monofilament mesh was attached to a 1 m high stainless steel frame with an arc and a chord length of 1.6 m and 1.4 m, respectively. The hydrophones were arranged in a gridlike fashion of five rows (three or six elements per row) and six columns (three or five elements per column), with a 27 mm horizontal and 18 mm vertical separation between each hydrophone (Fig. 3). Thus, utilizing the source of the echolocation clicks as a reference point, a 12° horizontal separation and 8.2° vertical separation existed between hydrophones. The arc of the steel frame was constructed such that each hydrophone was equidistant (1.2 m) from the dolphin's presumed anatomical source for echolocation click production.

Each hydrophone was coupled to an analog in-line filter-amplifier (Reson VP 1000) set for 20 dB of gain and band-passed from 10 to 200 kHz. For each of the 24 hydrophones, recorded echolocation clicks were digitized and stored to a PC using a 32-channel Interactive Circuits and Systems Limited analog to digital converter (312.5 kHz sample rate per channel, 16 bit resolution, 16 Mbytes/s transfer rate). Data were collected for a period of 5 s from the start of each trial, which was sufficient time for the dolphin to echolocate and make a decision regarding the presence of the target. The maximum peak pressure level capable of being recording was 193 dB re 1 μPa (i.e., saturation occurred at levels above 193 dB). Given the relatively close proximity and the good TSs, this limit was suitable for the recording of all echolocation clicks.

In the training sessions prior to data collection, various adjustments were made to the angle of the biteplate until the animal's head position was such that the MRA of his beam appeared on a central hydrophone of the hydrophone array. During echolocation data collection, the animals' head was immobilized using a biteplate which fixed the head and aligned the longitudinal axis of the animal's head with the center hydrophone and initial location of target presentation. The axis from the animal, through the center hydrophone, and to the target, was designated the 0° *position* (P_0; Fig. 3). Baseline data were collected with the targets on P_0 and then they were moved laterally, left or right, along a platform suspended across the pen. As targets were moved to further displacements, the target range increased. The distance from the dolphin's melon to the target at P_0 was 9.1 m; at the maximum lateral displacement of 5.0 m (28.8°), the distance to the target increased to 10.4 m (a difference of 1.3 m relative to the P_0 placement and a difference in transmission loss of ~2 dB).

C. Behavioral procedures

The dolphin was cued by the trainer to station on the biteplate at the beginning of each trial and to await the removal of the acoustically opaque screen. The time between biteplate stationing and screen removal was ~10 s. During this waiting period, the target was gently placed into position for target present trials. Target inspection was initiated by raising the acoustically opaque screen. The screen remained raised for 5 s, during which time the dolphin echolocated from his fixed position. Vision remained blocked during echolocation by the visually opaque but acoustically transparent polyethylene barrier. A go/no-go response paradigm was used for reporting target condition. The dolphin reported the target "present" by leaving the biteplate within 5 s of the trial start and touching a nearby paddle with his rostrum or "absent" by remaining on the biteplate for the 5 s trial duration. Correct reporting was reinforced with a fish reward, and no reinforcement was provided for incorrect reports. An equal number of target present and absent trials was conducted each session. Trials were presented in a random order according to a modification of the Gellerman series (Gellerman, 1933); i.e., the first-order conditional probability of a target present or absent trial was 0.50 and the second-order conditional probability of a ten-trial block beginning with a target present or absent was also 0.50.

Angular detection thresholds were estimated using an up/down staircase titration method (Moore and Schusterman, 1987) in which the target was successively moved further away from P_0 in one direction (left or right). Changes in the angular distance for each target present trial were determined by the dolphin's performance on the previous present trial. Distance was increased (30.48 cm) after a correct detection and decreased (30.48 cm) whenever the animal failed to report the presence of the target (a miss).

D. Data analysis

The average angle after 10–12 directional changes was used to define the angular detection threshold (i.e., the 50% correct detection performance). One threshold was calculated for both the SPH and the CYL to the right and left, resulting in a total of four thresholds. For the CYL, angular detection thresholds to the left and right of P_0 were determined with ten reversals. For the SPH, angular detection thresholds to the left were determined with ten reversals while those to the right were determined from two sessions of six reversals each (i.e., the two sessions of six reversals each were combined to determine the final threshold estimate). False alarm rates were calculated for both targets for displacements to the right and left of P_0.

Data collection and analysis were accomplished using custom designed analysis and visualization tools in LABVIEW 7 EXPRESS©. For each trial, the data stream recorded at the P_0 hydrophone was processed to detect and extract echolocation clicks. An emitted signal peak was considered a click when the instantaneous peak pressure of the input signal exceeded a threshold of 163 dB re 1 μPa. Detected echolocation clicks were recorded as a 128-point waveform, with a 32-point buffer prior to the pressure peak. Following data collection, extracted signals were visually inspected to verify that they were clicks and not transient signals due to other sources in the environment. If a click was detected at the P_0 hydrophone, the corresponding time-synchronized data were extracted from each of the other hydrophones in the array. Each click from each hydrophone was transformed to the spectral domain by applying a 64-point fast Fourier transform, without windowing, thus producing a frequency spectra representation with a 2.4 kHz bin resolution. Peak frequency, instantaneous peak pressure, center frequency, and rms bandwidth were calculated for each echolocation click.

The MRA of the echolocation beam was determined as the angle corresponding to the hydrophone that recorded the maximum level for each click. Horizontal and vertical −3 dB beamwidths were subsequently calculated assuming the MRA hydrophone as the center of the echolocation beam. Linear interpolation was used to determine the angular offset to the left and right of the MRA hydrophone that corresponded to the amplitude of 3 dB below the peak amplitude of the click recorded on the MRA hydrophone. Beamwidths were only calculated under the condition that the MRA hydrophone was not a peripheral element of the hydrophone array, i.e., the MRA hydrophone had to be in-line with no less than two additional hydrophones horizontally or vertically, depending on the beam to be measured. In addition, the vertical or horizontal in-line hydrophones to either side of the MRA hydrophone had to minimally span a click peak level of −3 dB of that recorded at the MRA hydrophone. If this criterion was not met, then beamwidth was not calculated for that click. Comparisons of horizontal and vertical beams within a click were restricted to matched samples, i.e., both the horizontal and vertical beamwidths were calculated on the same click. However, for average beam pattern measurements collapsed across conditions, all calculated beamwidths were taken into account.

Correlations between click characteristics (peak frequency, center frequency, peak level, horizontal beamwidth, vertical beamwidth, and rms bandwidth) were run for all clicks produced in target absent and present trials. Differences in click characteristics as a function of target type and position were assessed with a two-factor analysis of variance (two target types × four target positions). Target types were SPH and CYL, and the target position consisted of quadrants defined as near right, near left, far right, and far left. Near quadrants were defined by data collected when the target was presented from 0° to 15° to the left or right of P_0 and far quadrants were defined by data collected when the target was presented >15° to the left or right of the P_0. This series of tests was not performed for measurements of the vertical beamwidth as the loss of functionality of a hydrophone on the array during a number of sessions prevented the calculation of the vertical beamwidth (note that $N=6$ for the "near left" position of the CYL, Table II). Differences between individual quadrants were assessed with Tukey's Honestly Significant Difference (HSD) *post hoc* test of significance. All statistical tests were conducted with $\alpha<0.05$.

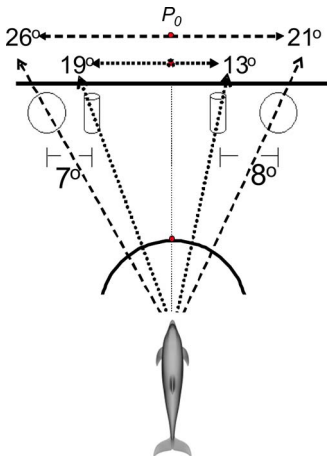


FIG. 4. (Color online) Right/left angular detection thresholds of the bottlenose dolphin for two target shapes. Detection thresholds for the cylinder (CYL) are associated with the dotted lines and those for the sphere (SPH) are associated with the dashed line. Angular detection thresholds were 7° (left) and 8° (right) greater for the SPH than the CYL.

III. RESULTS

A. Angular detection thresholds

A grand total of 149 target present trials, ranging from 2° to 29° to the left and from 2° to 21° to the right of P_0, was conducted during threshold testing. Angular detection thresholds are shown in Fig. 4. Angular detection thresholds for the SPH were greater to both the right and left than those for the CYL. The SPH detection threshold was 26° to the left of P_0 (FA=0.17) and 21° to the right (FA=0.12). The CYL detection threshold was 19° to the left of P_0 (FA=0.15) and 13° to the right (FA=0.19). Differences between SPH and CYL angular detection thresholds were relatively equal for the left (7°) and right (8°) target displacements.

B. Click production

The dolphin emitted an average of 36 clicks per trial for all trials conducted during threshold sessions. Table I gives the mean and standard deviation of the peak frequency, center frequency, rms bandwidth, and peak instantaneous sound pressure level for clicks emitted during target present trials. Significantly more clicks were produced when targets were presented in the far quadrants than when they were presented in the near quadrant [$F(3,139)=12.985$, $p<0.001$]. However, the target type had no effect on the number of clicks produced (Fig. 5). Target position showed a main effect with click peak pressures being significantly lower when the target was placed in the near left quadrant [$F(3,5438)=12.502$, $p<0.001$; Fig. 6]. In addition, no effect of target type was noted on the instantaneous peak pressure of clicks emitted during target present trials.

TABLE I. Characteristics of echolocation clicks produced during the threshold sessions. The degree of target displacement from P_0 corresponds to the following: near=0°–15°, far>15°. Values are means ±SEM.

Target	Quadrant	Peak frequency (kHz)	Center frequency (kHz)	rms bandwidth (kHz)	Peak SPL (dB at 1 μPa)	*N*
CYL	Near left	99.5±1.8	62.0±0.5	31.5±0.3	185.0±0.4	255
	Far left	97.0±0.9	61.7±0.2	30.4±0.1	187.1±0.2	1086
	Near right	75.0±0.8	66.5±0.2	34.2+0.1	187.4±0.2	1339
	Far right	73.7±3.4	64.7±0.9	30.8±0.6	187.4±0.8	69
SPH	Near left	88.7±2.2	62.0±0.6	31.6±0.4	184.7±0.6	157
	Far left	80.0±0.9	59.8±0.3	29.1±0.2	187.2±0.2	919
	Near right	101.0±1.3	60.1±0.4	29.9±0.2	186.3±0.3	449
	Far right	87.6±0.8	62.0±0.2	31.0±0.1	187.1±0.2	1172

C. Beamwidths and the beam MRA

A composite beam pattern measured when the target was directly in front of the dolphin is presented in Fig. 7. A representative beam pattern collected when the target was presented far to the left or right of P_0 is also presented. A composite beam pattern for targets displaced to the left or right of P_0 is not presented as the dolphin's search strategy involved a combination of forward looking narrower beams as well as those steered and widened in response to target placement. Thus, a composite beam does not adequately depict the maximal beam control capabilities of the dolphin. Table II gives the mean and standard deviation of beamwidths according to target type and target location. Across all trials, the horizontal −3 dB beamwidth ranged from 15° to 40°, averaged 21°, and had a bimodal distribution with beamwidths clustering about 16° and 26° (Fig. 8). The vertical −3 dB beamwidth ranged from 10° to 29° and averaged 18°. For both the vertical and horizontal beams, beamwidths were typically less than 28°. (It is important to note that the beamwidth estimates are estimates of the maximum beamwidth. The spatial resolution of the recording hydrophones results in an ambiguity in the actual beamwidth. Based on a symmetrical loss at the midpoint of two hydrophones, beamwidths may be overestimated by as much as 6° in the horizontal plane and 4.1° in the vertical plane. Thus, the range of horizontal −3 dB beamwidths measured could be reduced to 9°–34° and from 4°–23° in the vertical.)

Horizontal beamwidths were significantly negatively correlated with peak frequencies ($r=-0.31$); significant relationships were observed with other click characteristics (center frequency, vertical beamwidth, and rms bandwidth); however, all these relationships were weakly correlated ($-0.04<r<0.07$). The vertical beamwidth was positively correlated with peak level ($r=0.15$). Significant relationships were observed with other click characteristics, but as with the horizontal beamwidth, these relationships were weak ($-0.04<r<0.07$); indeed, the vertical and horizontal beamwidths varied nearly independent of one another ($r=-0.04$).

Figure 9 shows the number of times each hydrophone in the 24 element array was on/closest to the MRA for each of the target present trials according to whether the SPH or CYL was presented to the left or right of P_0. The hydrophone most closely corresponding to the MRA was typically the hydrophone at P_0 (stippled bar). However, the hydrophone most closely corresponding to the MRA was found to occur up to 16° and 24° away from P_0 in both the vertical and horizontal planes, respectively. The distribution of the MRA was notably skewed to the right when targets were presented to the right of P_0.

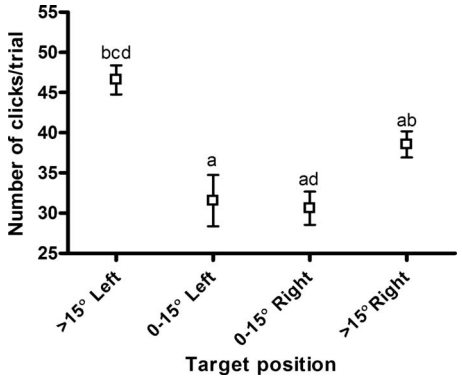


FIG. 5. Change in the average number of clicks emitted per trial as a function of target displacement from P_0. Lowercase letters over a displacement category indicate that a significant difference exists between it and another category [(a) ">15° left," (b) "0°–15° left," (c) "0°–15° right," and (d) ">15° right"].

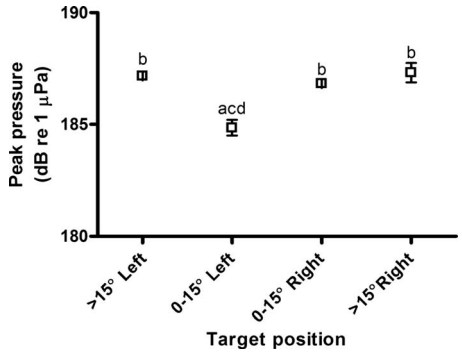


FIG. 6. Change in the average peak level of clicks emitted per trial as a function of target displacement from P_0. Lowercase letters over a displacement category indicate that a significant difference exists between it and another category [(a) ">15° left," (b) "0°–15° left," (c) "0°–15° right," and (d) ">15° right"].

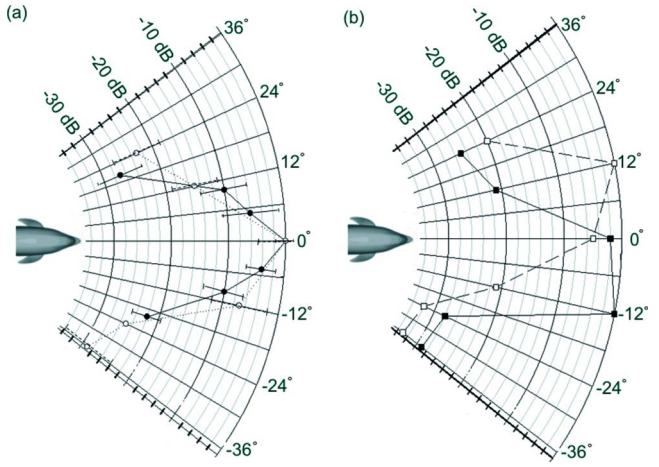


FIG. 7. (Color online) (a) Composite horizontal beam pattern from the bottlenose dolphin for targets presented directly in front of him (at P_0). Open circles correspond to the composite beam recorded in this study; filled circles correspond to the composite beam recorded by Au *et al.* (1986) from the same animal. (b) Representative beams for when the target was placed far to the left (open squares) or far to the right (filled squares) of P_0.

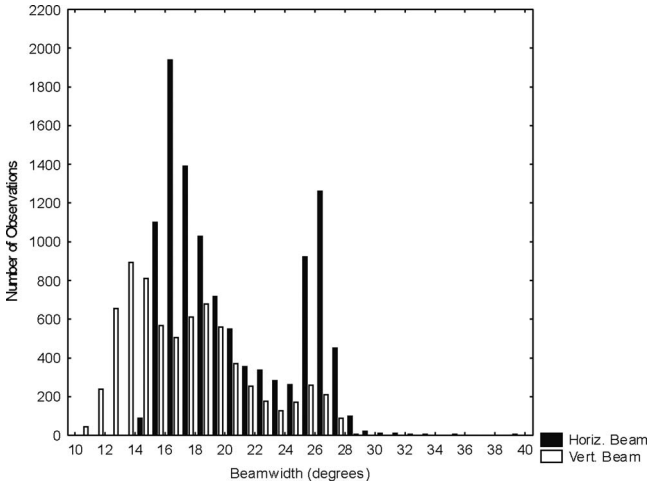


FIG. 8. Distribution of the vertical and horizontal beamwidths measured throughout all threshold session trials. Note the bimodal nature of the horizontal beamwidth distribution (between 24° and 28°).

IV. DISCUSSION

The most remarkable finding of this study was that the horizontal and vertical beamwidths of the echolocating dolphin were substantially broader than the composite beamwidths previously reported (Au *et al.*, 1978, 1986). Beam patterns of dolphins (*Tursiops truncatus*) have been studied by several investigators beginning with observations of target detection as a function of location about the head (Norris *et al.*, 1961) and initial observations of the form of the emitted echolocation beam (Evans, 1973). Evans (1973) was the first to attempt to measure echolocation beam patterns for individual clicks in a click train produced by a bottlenose dolphin. Using a ten-element vertical array of hydrophones 1.0 m from the assumed click source, Evans (1973) noted that dolphin beam patterns for individual clicks were "chaotic" and concluded that internal manipulation of the signal source was evident. A later study required a dolphin to station its head to ensure immobility, thus enabling more accurate measures of the beam MRA (Au *et al.*, 1986). Beam patterns reported from that study had composite horizontal and vertical −3 dB beamwidths of approximately 10°, with some degree of variability.

The dolphin participating in this study is the same dolphin used by Au *et al.* (1986). The data collected here indicate that in the angular detection of targets, this same dolphin produced horizontal and vertical echolocation beams up to 30° and 19° broader, respectively, than the composite beams previously reported. In the work of Au *et al.* (1986), the subject was tasked to echo detect a small steel sphere suspended 6.4 m directly in front of where the dolphin was stationed. It is likely that the dolphin settled on an effective potentially stereotypical beam pattern that did not move appreciably either horizontally or vertically. The dolphin would have had no need to grossly vary the emitted echolocation beam due to a predictable target location. Not surprisingly, the composite beam for clicks produced when the target was placed directly in front of the dolphin in this study is similar to those previously observed by Au *et al.* (1986).

Angular detection thresholds were not symmetrical about P_0. Detection thresholds for both targets to the right were reduced when compared to detection thresholds to the left of P_0. The most probable reason for the dolphin's asym-

TABLE II. Horizontal and vertical beamwidths of the dolphin echolocation beam measured during target (CYL and SPH) presentations to the right and left of P_0. The degree of target displacement from P_0 corresponds to the following: near=0°–15°, far>15°. Beamwidth values are means ±SEM.

Target	Quadrant	Horizontal beamwidth (deg)	*N*	Vertical beamwidth (deg)	*N*
CYL	Near left	18.4±0.3	240	15.3±1.7	6
	Far left	19.8±0.1	1038	17.2±0.4	93
	Near right	22.0±0.1	1328	17.8±0.1	1314
	Far right	22.9±0.5	68	16.5±0.5	68
SPH	Near left	21.0±0.3	149	17.7±0.3	148
	Far left	22.1±0.1	897	18.6±0.1	889
	Near right	18.9±0.2	445	16.1±0.2	338
	Far right	21.9±0.1	1093	17.8±0.1	805

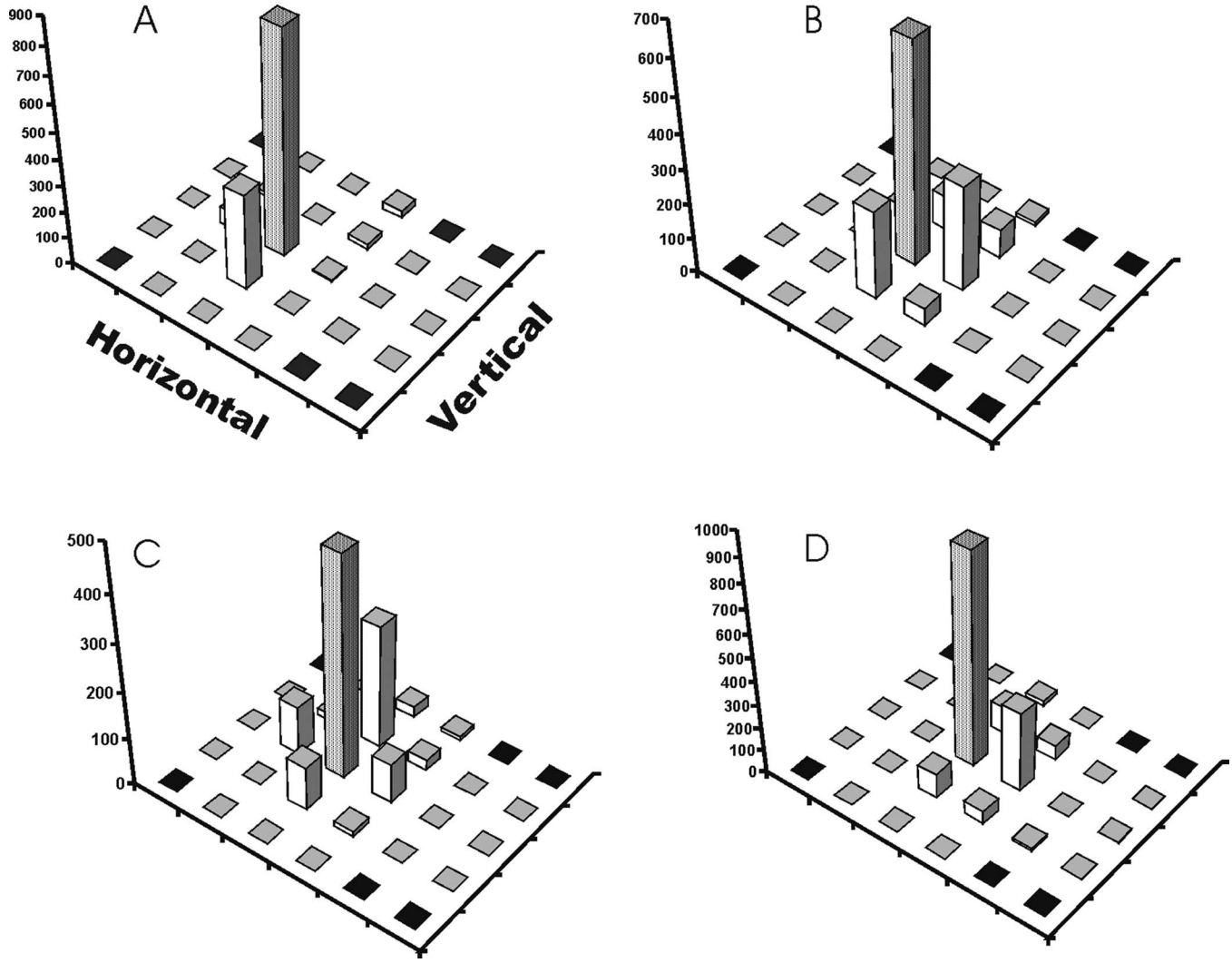


FIG. 9. Distribution of the hydrophones that registered echolocation clicks as being on-MRA during threshold sessions (bars denote number of clicks). Gray squares correspond to the location of hydrophones within the array (see Fig. 3 for a reference of hydrophone location); black squares correspond to empty space. The hydrophones are separated by 12° in the horizontal and 8.2° in the vertical. The distributions correspond to the following test scenarios: (a) CYL target displaced to the left, (b) CYL displaced right, (c) SPH displaced left, and (d) SPH displaced right.

metric detection thresholds is hearing loss. This animal is known to have hearing loss in both ears at frequencies greater than 40 kHz, and a 10–20 dB greater loss in the right ear at frequencies between 10 and 40 kHz (Brill *et al.*, 2001). Based on the measurement of TS for the two targets, the CYL should have been detected further from P_0 than the SPH for both right and left displacements. However, the angular detection threshold for the CYL is consistently less than that of the SPH. Inspection of echoes returning from the CYL during those trials indicated that the CYL did not always produce the higher amplitude echo. Due to the suspension method for the CYL, currents and wave action within the bay changed the aspect of the target. The TS of a cylinder is sensitive to orientation along its long axis, and movements away from vertical will change the effective length of the long axis and act to reduce the TS (Urick, 1983). Since the TS of a sphere is aspect independent, movement of the SPH due to currents and wave action did not appreciably affect the target's TS.

The clustering of horizontal beamwidths around 26° (Fig. 8) is strongly suggestive of the dolphin's attempt to broaden the area of ensonification when targets were presented off-MRA of the echolocation beam. Beamwidth variation is not a new finding in odontocete echolocation; in addition to the previous studies noting the variability in dolphin echolocation beam patterns (Evans, 1973; Au *et al.*, 1986), the beluga has also been reported to produce variable beam patterns in both the vertical and horizontal planes (Au *et al.*, 1987). However, this is the first quantification of the extent to which echolocation beamwidths are varied by a bottlenose dolphin participating in a target detection task where the targets are progressively displaced toward the edge of the echolocation beam. The distributional limits of the horizontal beamwidths, taken in concert with the angular detection thresholds, suggest that the maximum horizontal beamwidth capable of being produced by a bottlenose dolphin is approximately 40°, with most broadened beams being approximately 26° (second peak of the bimodally distributed horizontal beamwidth data; Fig. 8). Vertical beamwidths were also found to be greater than previously reported (up to 29°), but whether these represent the maximal extent of vertical beam broadening is unknown because the animal was not

tasked with detecting targets displaced in the vertical dimension. For either measurement, it must be acknowledged that the spatial resolution of the hydrophone array may result in an overestimate of the beamwidths by as much as 6° and 4.1° in the horizontal and vertical planes, respectively.

Assuming that the attenuation in receive level between adjacent hydrophones is linear and symmetrical in space, the midpoint between the MRA hydrophone and the vertical or horizontal hydrophone that approaches P_0 can be used to estimate a nominal beam steering ability in the vertical and horizontal planes. The nominal beam steering ability of the dolphin in this study was calculated to be 18° in the horizontal, 12° in the upward vertical, and 4° in the downward vertical. The subject demonstrated a strong right-sided bias in beam steering when targets were presented to the right of P_0. This finding is of interest because of the documented hearing loss within this animal which is greatest in the right ear (Brill *et al.*, 2001), a phenomenon which, as previously noted, also seems to be reflected in the smaller angular detection threshold to the right of the animal. Although the right-sided bias might be due to asymmetry of the dolphin skull, a comparison to the distribution of beamwidths when target displacement was to the left suggests that the dolphin may have actively compensated for its hearing loss in the right ear.

A complicated series of premaxillary and vestibular air sacs associated with the nasal passages and click apparatus as well as a fibrous sheath and underlying muscle associated with the melon (Litchfield *et al.*, 1979) all have the potential to contribute to variation in the emitted sonar beam. The air sacs can be inflated and deflated by the dolphin, and in the inflated state the air sacs should be useful as reflectors. That these variably reflective surfaces might contribute to beamforming has been previously proposed based on the modeling of beam formation via numerical integration of the wave equation (Aroyan *et al.*, 1992). Muscular deformation of the melon may also contribute to beamforming by varying the geometry of the melon or constricting its volume such that the density of the gradient within the melon changes. Melon deformations are not unheard of within the odontocetes, as have been noted by Norris for the beluga (Norris, 1968). In this study, slight deformations of the melon were observed as a melon "squeeze," suggesting that subtle variations in melon geometry/lipid density were potentially performed by the dolphin. Another intriguing possibility is that the dolphin phase shifts the two phonic lips in order to accomplish some nominal degree of beam steering. Utilization of certain joint time-frequency (JTF) analyses may permit this latter hypothesis to be tested; prior JTF analyses, utilizing the fractional Fourier transform, of dolphin clicks have demonstrated two separable chirplike components with the higher frequency component slightly delayed from 5 to 20 μs relative to the lower frequency component (Capus *et al.*, 2007). These findings are consistent with the proposition of two sources for click generation and support the idea that the sources might be time delayed during pulse generation to produce phase shifting.

The overall increase in click peak pressure for all targets displaced to the right of P_0 (Fig. 6), the predominant right-sided bias in the echolocation beam MRA (Fig. 9), and the occurrence of broader horizontal beams during angular target detections suggest that the dolphin consciously increased its effort in the detection of targets displaced to the right and when targets were placed further from P_0. The conscious ability to vary echolocation beamwidths and to steer the echolocation beam is an important consideration in determining how dolphins utilize echolocation. Past echolocation research has predominantly relied on experimental designs that impose artificiality in favor of control over the recording of echolocation (Au *et al.*, 1995, 1999; Houser *et al.*, 1999). Because stereotypical echolocation strategies may result from these designs, the richness and variability in echolocation click production as a function of task and environment are likely obscured.

In studies of echolocation in natural or free-swimming settings, large samples of on-axis clicks can be difficult to obtain, and assumptions about their properties may become necessary in order to estimate when clicks are on-MRA (Au and Herzing, 2003; Møhl *et al.*, 2003; Philips *et al.*, 2003; Au *et al.*, 2004; Zimmer *et al.*, 2005). Both the potential bias introduced in the subsampling of clicks and a smaller sample size can hinder analyses of how the physical characteristics of the clicks (e.g., peak pressure and spectral content) vary as a function of the task and may make it difficult to accurately describe click structure and variability. Descriptions of echolocation beamwidths in free-swimming studies likewise rely on relatively few recordings of clicks estimated to be on-MRA and which are recorded at relatively coarse spatial resolutions (Rasmussen *et al.*, 2004; Zimmer *et al.*, 2005). As noted by Madsen and Wahlberg (2007), the difficulties inherent in field measurements provide a number of potential pitfalls to the accurate characterization of echolocation clicks. If wild odontocetes are beam steering or varying their beam patterns at the time of data collection, then the difficulties are potentially amplified. The contrast in results from this study to those of free-swimming odontocetes illustrates the need for consideration of experimental design and method limitations when interpreting the nature of echolocation beamwidth and click characteristics.

V. CONCLUSIONS

Echolocation beam geometry and beam steering are plastic and are consciously controlled by the animal. The echolocation beamwidth and direction of steering are varied by the animal based on task requirements. Estimated maximum horizontal beamwidths up to 40° and estimated vertical beamwidths of up to 20° can be produced by the bottlenose dolphin. Targets presented at greater peripheral distances from the dolphin require greater effort (additional clicks/higher peak pressure) to be detected. Two mechanisms are proposed as possible means of beamwidth modulation/steering: (1) phase shifting between click generators and/or (2) manipulation of the air-sac/melon volume and geometry.

ACKNOWLEDGMENTS

The authors would like to thank M. Cross for the machine code programming of the DAQ system, S. Martin for designing the binaural receiving array, D. Kastak for helpful

comments during the course of the experiment, and B. Branstetter for commenting on the paper. The authors would also like to thank E. Testa and J. Bridger for training support, and participants of the marine mammal internship program of SSC SD for additional logistical support. This project was funded by the Office of Naval Research (Funding No. N0001403WX21053). All research procedures were conducted in accordance with the Ethical Principles of the Acoustical Society of America for Research Involving Human and Non-Human Animals in Research and Publishing and Presentations as established by the Acoustical Society of America.

Akamatsu, T., Wang, D., and Wang, K. (**2005**). "Off-axis sonar beam pattern of free-ranging finless porpoises measured by a stereo pulse event data logger," J. Acoust. Soc. Am. **117**, 3325–3330.

Aroyan, J. L., Cranford, T. W., Kent, J., and Norris, K. S. (**1992**). "Computer modeling of acoustic beam formation in *Delphinus delphis*," J. Acoust. Soc. Am. **92**, 2539–2545.

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. L., and Snyder, K. J. (**1980**). "Long-range target detection in open waters by an echolocating Atlantic bottlenosed dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **68**, 1077–1084.

Au, W. W. L. (**1980**). in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 251–282.

Au, W. W. L., Floyd, R. W., and Haun, J. E. (**1978**). "Propagation of Atlantic bottlenose dolphin echolocation signals," J. Acoust. Soc. Am. **64**, 411–422.

Au, W. W. L., Kastelein, R. A., Rippe, T., and Schooneman, N. M. (**1999**)."Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **106**, 3699–3705.

Au, W. W. L., Moore, P. W. B., and Pawloski, D. (**1986**). "Echolocation transmitting beam of the Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **80**, 688–694

Au, W. W. L., and Pawloski, D. A. (**1992**). "Cylinder wall thickness difference discrimination by an echolocating Atlantic bottlenose dolphin," J. Comp. Physiol. [A] **170**, 41–47.

Au, W. W. L., Pawloski, J., Nachtigall, P. E., Blonz, M., and Gisner, R. C. (**1995**). "Echolocation signals and transmission beam pattern of a false killer whale (*Pseudorca crassidens*)," J. Acoust. Soc. Am. **98**, 51–59.

Au, W. W. L., Penner, R. H., and Turl, C. W. (**1987**). "Propagation of beluga echolocation signals," J. Acoust. Soc. Am. **82**, 807–813.

Au, W. W. L., Ford, J. K. B., Horne, J. K., and Newman Allman, K. A. (**2004**). "Echolocation signals of free-ranging killer whales (*Orcinus orca*) and modeling of foraging for Chinook salmon (*Oncorhynchus tshawytscha*)," J. Acoust. Soc. Am. **115**, 901–909.

Brill, R. L., Moore, P. W. B., and Dankiewicz, L. A. (**2001**). "Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones," J. Acoust. Soc. Am. **109**, 1717–1722.

Capus, C., Pailhas, Y., Brown, K., Lane, D. M., Moore, P. W., and Houser, D. (**2007**). "Bio-inspired wideband sonar signals based on observations of the bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **121**, 594–604.

Cranford, T. W. (**1992**). "Functional morphology of the odontocete forehead: Implications for sound generation," Ph.D. thesis, University of California, Santa Cruz, CA.

Cranford, T. W. (**2000**). in *Hearing by Whales and Dolphins*, edited by W. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 109–156.

Cranford, T. W., Amundin, M., and Norris, K. S. (**1996**). "Functional morphology and homology in the odontocete nasal complex-implications for sound generation," J. Morphol. **228**, 223–285.

Evans, W. E. (**1973**). "Echolocation by marine delphinids and one species of fresh-water dolphin," J. Acoust. Soc. Am. **54**, 191–199.

Gellerman, L. W. (**1933**). "Chance orders of alternating stimuli in visual discrimination experiments," J. Genet. Psychol. **42**, 206–208.

Hammer, C. E., Jr., and Au, W. W. L. (**1980**). "Porpoise echo-recognition: An analysis of controlling target characteristics," J. Acoust. Soc. Am. **68**, 1285–1293.

Houser, D. S., Helweg, D. A., and Moore, P. W. B. (**1999**). "Classification of dolphin echolocation clicks by energy and frequency distributions," J. Acoust. Soc. Am. **106**, 1579–1585.

Houser, D. S., Martin, S. W., Bauer, E. J., Phillips, M., Herrin, T., Cross, M., Vidal, A., and Moore, P. W. (**2005**). "Echolocation characteristics of free-swimming bottlenose dolphins during object detection and identification," J. Acoust. Soc. Am. **117**, 2308–2317.

Litchfield, C., Karol, R., Mullen, M. E., Dilger, J. P., and Luthi, B. (**1979**). "Physical factors influencing refraction of the echolocative sound beam in delphinid cetaceans," Mar. Biol. (Berlin) **52**, 285–290.

Madsen, P. T. and Wahlberg, M. (**2007**). "Recording and quantification of ultrasonic echolocation clicks from free-ranging toothed whales," Deep-Sea Res., Part I **54**, 1421–1444.

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

Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2003**). "The monopulsed nature of sperm whale clicks," J. Acoust. Soc. Am. **114**, 1143–1154.

Moore, P. W., and Schusterman, R. J. (**1987**). "Audiometric assessment of northern fur seals, *Callorhinus ursinus*," Marine Mammal Sci. **3**, 31–53.

Nachtigall, P. E. (**1980**). in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 71–95.

Norris, K. S. (**1968**). in *Evolution and Environment*, edited by E. T. Drake (Yale University Press, New Haven, CT), pp. 297–324.

Philips, J. D., Nachtigall, P. E., Au, W. W. L., Pawloski, J. L., and Roitblat, H. L. (**2003**). "Echolocation in the Risso's dolphin, *Grampus griseus*," J. Acoust. Soc. Am. **113**, 605–616.

Norris, K. S., Prescott, J. H., Asa-Dorian, P. V., and Perkins, P. (**1961**). "An experimental demonstration of echolocation behavior in the porpoise *Tursiops truncatus* (Montagu)," Biol. Bull. **120**, 163–176.

Pilleri, G. (**1982**). in *Investigations on Cetacea*, edited by G. Pilleri (University of Bern, Berne, Switzerland), Vol. **XIII**, pp. 167–176.

Rasmussen, M. H., Wahlberg, M., and Miller, L. A. (**2004**). "Estimated transmission beam pattern of clicks recorded from free-ranging white-beaked dolphins (*Lagenorhynchus albirostris*)," J. Acoust. Soc. Am. **116**, 1826–1831.

Urick, R. J. (**1983**). *Principles of Underwater Sound* (Peninsula, Los Altos, CA).

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.